Author response to Anonymous Reviewer #1 on: "The NASA Eulerian Snow on Sea Ice Model (NESOSIM): Initial model development and analysis" ***by*** **Alek A. Petty et al.**

Reviewer comments are in black, our responses are in blue.

We will also submit the revised manuscript and a word document highlighting the tracked changes we have made based on these comments.

The paper describes the development, calibration and validation, and sensitivity of a snow on sea ice model. Given that the model is likely to be used to retrieve ice thickness from CryoSat and ICESat altimetry, the model should be documented in the literature. However, the paper needs to be improved and some points clarified before it can be published. I would suggest that the authors consider reorganizing the paper to make it more readable and potentially shorter.

We sincerely thank the reviewer for providing this review. It included some very helpful suggestions on how we can improve the manuscript, and areas where we need to provide more justification etc.

General Comments

While I recognize that a model cannot include everything, I am surprised that the authors do not include snow melt. I see this as a major shortcoming in the model. The top-left panel of Figure 3 shows, what might be interpreted, as a melt signal between March and April. I suspect that melt is also a factor in the North Atlantic sector. A warming Arctic is almost certainly likely to have melt earlier. This years warming 'spike' likely caused melting in some sections of the Arctic. While this might not have resulted in a loss of snow mass, it would have increased snow density and caused a reduction in snow depth - refreezing/metamorphosis is another process. The authors should address in more detail leaving these key processes out.

We have added more details in the summary of planned future model improvements, including the lack of snow thermodynamics (highest priority). We feel this discussion combined with the comments in the introduction make clear we see the lack of melt processes as the main shortcoming of the model.

We have also added the following to the end of the abstract: ' Potential improvements to this initial NESOSIM formulation are discussed in the hopes of improving the accuracy and reliability of these simulated snow depth and density.'

Regarding the decrease in the observed snow depths in the Soviet Station data: we actually doubt this is due to a melt event. Air temperatures remained below freezing for all of April for each station in 1980-1991. None of the stations showed a decrease in snow depth except for 1 (Station 31) in 1990 – the air temperatures rose to -4C, which could potentially have been a melt event given that it's a daily average. Looking at the sea level pressure for that time, there was a big decrease between April 10 and April 30 (the time when the decrease in snow depth occurs) – which suggests there may have been a storm event, which likely redistributed the snow and may have contributed to a decrease in mean depth along the survey line. A storm event would also explain the decrease in density (e.g., fresh snowfall).

Another point to consider is that we only took data from four stations (1980-1991). Despite some of them lasting for multiple years, this is still a relatively small sample size, and the mean in snow depth will be strongly affected by variations in a single data point. We refer the reviewer to

We have also added a lot more justification at the start and end of the revised manuscript regarding our approach and expectations for future model development/calibration efforts to improve these error.s

We have also added the following to the end of the abstract: ' Potential improvements to this initial NESOSIM formulation are discussed in the hopes of improving the accuracy and reliability in these simulated snow depth and density'

The development of a snow product for improving retrieval of sea ice thickness from altimetry is critical for ICESAT 2 to be useful and this team should have NASA's support to do just that. Such a snow model's accuracy goal must be based on a desired accuracy in thickness retrievals. (e.g. retrieval of ice thickness accurate to +-0.5m over a given domain demands snow depth accurate to O5cm over the same domain). The model presented here is not up to meeting these kinds of needs, and does not leverage the existing (more sophisticated) models of snow on sea ice (e.g. LIM, SnowModel, CICE).

We are not aware of more sophisticated approaches providing the requirements you list, and our belief is that uncertainties in the input forcing data needed to be explored further, along with efforts to improve model sophistication. We see this model as a contribution towards this end goal. Based on the comments of the other reviewers we have added more discussion of the other snow models available and the physics they include.

We agree that working towards a 5 cm error makes sense, and we believe this provides a useful framework to build towards that goal. This is also discussed in a new study by a co-author of this study: Webster et al. (accepted) which speaks to several issues of our current abilities in observing and treating snow in models. Existing basin-scale observations unfortunately do not have 5-cm accuracy; snow depths derived from merged satellite data do not have 5-cm accuracy. So it is difficult to produce model output with 5-cm accuracy given that no observations of 5-cm accuracy (or better) at the basin-scale exist for model validation and assessment, and especially on a seasonal time-scale.

Some major issues include: Model design relative to state of knowledge: 1. The two layers used in the two layer model (new snow and windslab) do not represent the two layers of the snowpack discussed in literature (wind slab and depth hoar). Authors cite and discuss the literature indicating that windslab and depth hoar dominate the mass of the pack and have quite different density – then ignore these decades of observation to invent a new scheme unsupported by observations. Respecting the effort to create a simple, 2 layer model, new snow should not be one of the two layers. The references cited clearly state that new snow rarely comprises much of the Arctic snowpack, because it is very rapidly converted to windslab. The preservation of a new snow layer appeared to be designed for modeling loss of snow into leads – but little is known about the magnitude of this flux, and it was minor in this model. 2. The model is operated on a 100x100km grid, which is very coarse relative to the variability in ice – which is shown to be important in impacting the accumulation of snow. The data sets used provide much higher ice concentration, and movement information – this data should be used at full resolution and atmospheric data can be downsampled. 3. Melt is neglected despite it being important during part of the timeframe and having significant impact on results.

1. We expanded on the description in the manuscript about the second layer representing depth hoar and wind slab and their respective densities. We treat wind slab and depth hoar by taking the average ratio of these two layers, based on Radionov et al. 1997 and Sturm et al. 2002, over the

Warren et al. (1999) which shows data from all stations (1954-1991) and thus is more representative of the seasonal cycle.

Another concern is the use of the Polar Stereographic grid. The model tracks snow volume but only sea ice concentration, and not area, appears in Equation 1. Are the authors assuming that Polar Stereographic grids are equal area? This is not the case, they are conformal but not equal area. Cells at 70 N are about 10% smaller than cells at the pole. Maybe I am missing something here, and maybe this erroneous assumption might not have a big impact, but the authors should satisfy both themselves and readers that the choice of grid does not have an impact.

This is an interesting point, and working with projections/grids always provide some complexity that should not be quickly overlooked. In this case we are using equal areas in our polar stereographic projection, so quantities are being conserved in the model. However the reviewer is correct that the polar stereographic projection does provide a distortion to the real world such that our grid cells will be larger at lower latitudes. We do not see this as a huge concern as this is similar to the issue one always has when using polar stereographic data. We did consider using equal area projection (the Lambert azimuthal projection) but this can introduce potential issues with the vector projections in this non-conformal projection that could offset the potential benefits of less distorted area grids.

An obvious question, given the prevalence of Warren 1999 snow depths in sea ice thickness retrievals, is how different are the results presented here from W99? I would argue that W99 is the current benchmark for evaluation of snow depth products. The authors should include some discussion on this topic. I think it would also be useful to put the uncertainty reported in this paper (~ 10 cm) in the context of thickness retrievals. Ten centimeters is at least 30% of snow depths in the Arctic and similar to inter-annual variability in snow depth. While there is clearly room for model improvements, quality of precipitation fields and other forcing fields also come into play. It would be good to discuss these issues.

Another good point. Instead of simply comparing the results against W99, we believe the more interesting comparison is between W99 (actually the modified W99) and OIB (ground truth) and NESOSIM and OIB. This was similar to the approach taken by Blanchard-Wigglesworth et al., (2018) in their recent model evaluation efforts (published during this discussion period).

We have carried out this analysis showing that the difference between the model and mW99 compared to OIB is similar, but also depends on the chosen OIB product. See the figure attached:

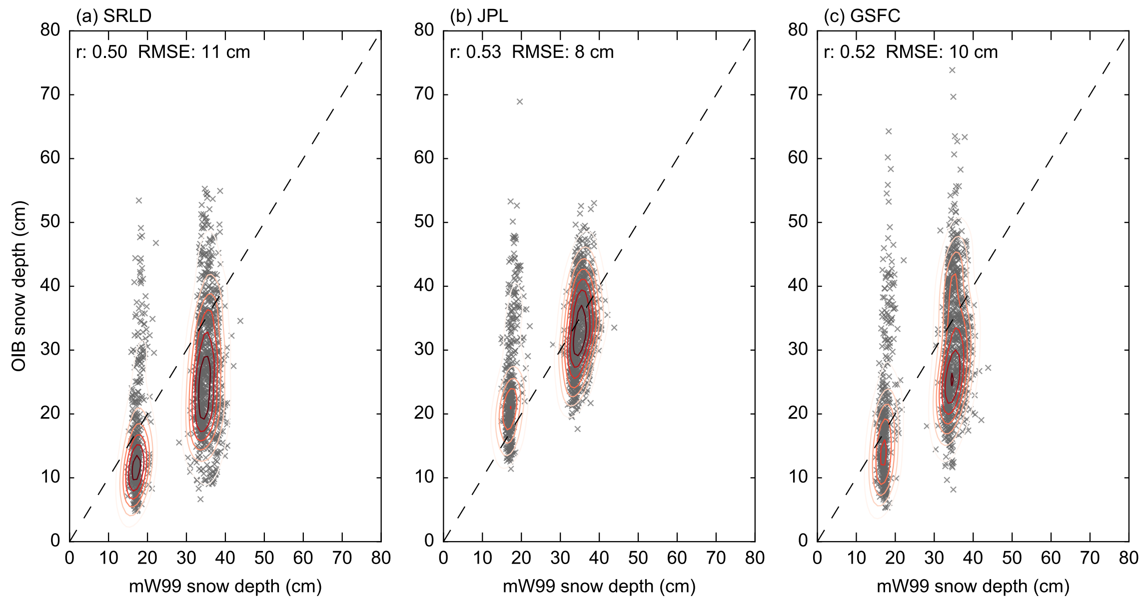


Figure 1: Comparison of modified Warren snow depths against OIB snow depths for the years 2009-2015. The data are binned to the 100 km model grid, only for grid-cells with significant (>1000 OIB data points).

While the model doesn't really offer an improvement over modified Warren in terms of these comparisons, our hope is that with continued model developments and calibrations we can improve the performance and reliability. Uncertainty in the OIB data and the possibility they have been tuned to fit the modified Warren climatology make this interpretation more challenging.

We have added more discussion of these comparisons, and the comparisons to mW99 to the revised manuscript.

I find the model formulation as described in section and equations 1 through 8 confusing. This may be because I think of the process in a different way to that described here. Hopefully, my interpretation in the following paragraph, whether right or wrong, will help improve the model description.

Yes we appreciate these efforts to make the model formulation more readable. As you say, we are really tracking an effective snow depth, and we have tried to make this clearer in the revised manuscript.

I see the model as analogous to the evolution equations for ice thickness (or any other tracer), where the change in snow depth is the sum of a dynamic component (-del*dot* (hu) [I think] not del(hu)), and a static "snow depth evolution" component that rep- resents snow accumulation, ablation by wind, and compaction (h^{acc}-h^{wp}+h^{bs} for the "new" snow layer and h^{wp} for the "old" snow layer). What I find confusing is that h^{acc} is the product of snow accumulation (snowfall/density) and sea ice concentration. Shouldn't what I call the sum representing the "snow depth evolution" component be multiplied by concentration (A) not just (Sf/density); e.g. A*(Sf/density -h^{wp}+h^{bs}) [Note my comment in the paragraph above – this only applies to a uniform grid]. Similarly, should h in –del *dot* (hu) be Ah.

I think our use of effective snow depth was confusing things. As you say we are really tracking Ah, which is why our use of A only really comes in when we introduce the accumulation term (if we keep using it we double count). The other terms then act on this effective snow depth, so we don't need to apply the ice concentration again in those terms.

Maybe this is what the authors mean by “we track snow volume” – e.g. (Ah) – and “An effective snow depth...”. However, I would suggest that it is Ah that is the effective snow depth because this term represents a mean grid cell snow depth (including open water areas). Whereas, h can be thought of as a physical snow depth because it represents the process of accumulation, wind ablation and compaction at point on an ice floe. I think what I describe is a conceptual difference rather than an error in the model because h^{wp} and h^{bs} are tuned, so the wind packing and blowing snow coefficients can be thought of as including sea ice concentration, i.e. you can change the model description in the paper without changing the model.

Agreed. We have changed this in the revised manuscript to read as: 'Note that in the model we track the evolution of an effective snow depth within each grid-cell (the volume of snow per unit grid cell area) for simplicity. The actual snow depth over the ice fraction is calculated by dividing the effective grid-cell snow depth by the grid-cell ice concentration.'

Two further issues are with model calibration and the use of the ensemble mean snowfall. If I understand correctly, the parameters are tuned using ERA-Interim snowfall and these “best” parameters are applied to model runs with MERRA, JRA55 and the Ensemble Snowfall. I would suggest this is the wrong approach. The calibration process will compensate for biases in ERA-Interim snowfall. However, biases in the other reanalyses are different. A different “best” parameter set should be expected for each reanalysis snowfall product. Conceptually, model equations, parameters and forcing data are all part of the Model. Using parameters obtained for ERA-Interim, might have detrimental effects on snow depths when other reanalyses are used. I would recommend that the MERRA, JRA55 and (maybe) ensemble runs should be calibrated separately.

We did consider this, and the idea makes sense. However, in reality the tuning was not highly optimized and instead we attempted to find parameters that achieved a good balance between capturing the seasonal cycles in snow depth and density in the Soviet Station data. Again, this was mostly due to concerns regarding how representative the Soviet Station data were for calibration purposes. We have added the following to the revised manuscript to explain this:

P17: "We also decided against specific model configuration parameter tuning due to the limitations in the calibration data, however this should be considered when analyzing the model performance, especially with regard to our validation efforts (i.e. more sophisticated and/or configuration specific tuning could improve the comparisons shown)."

This was also discussed in Section 4: " Specific model configurations may be required based on user demands, and our expectations is for these calibrations to evolve as new calibration data are made available and physical parameterizations introduced/updated."

We also added the following to this paragraph: " As discussed in Section 3, it is likely that specific model configuration tuning could improve these comparisons and the later validation efforts, but we decided against a more optimized calibration approach due to the limitations in the Soviet station data."

With regards to the Ensemble snowfall, the assumption behind an ensemble average being a better estimate is that individual ensemble members bracket “reality”. Is this the case with reanalysis snowfall? If all ensemble members are biased in one direction, the ensemble average will also be biased. My understanding is that both ERA-Interim and MERRA precipitation are both biased high compared to land stations. Based on Fig 11, JRA55 is also high. So is the ensemble snowfall an improvement over the individual ensemble members?

This is a good point and one we perhaps didn't make clear enough in our discussion of the ensemble 'median' snowfall data. We have dropped the idea that the median snowfall dataset might be somehow 'better' from Section 4 ("as we expect these results to be less prone to errors in the individual reanalyses etc.), so this now reads:
"we choose to mainly focus on the MEDIAN-SF forced results using the default configuration (Table 1) for simplicity"

We don't feel that land stations are wholly representative of conditions over the sea ice (confirmed from conversations with colleagues in NASA's GMAO for example) hence our recent study currently in press looking at precipitation estimates over the Arctic Ocean from reanalyses (Boisvert et al., 2018). While again we are limited in direct observations over sea ice, comparisons of the precipitation converted to snow depth (using constant snow density approximations and lagrangian feature tracking) against drifting snow buoys in the central Arctic Ocean showed no obvious bias in the differences between the reanalysis derived snow depth estimates used in this study and the buoy snow depths (in Boisvert et al., 2018). It did appear that MERRA-2 has a clear positive bias so was dropped from our analysis. Clearly more work needs to be done to better understand the precip within the Central Arctic, however.

There needs to be more detail about the ensemble snowfall was generated. For example, I can envisage ERA-Interim, MERRA and JRA55 all having snow but the location of this event being shifted by one or two grid cells. On one side of the event, while ERA-Interim might have no snow, MERRA and JRA55 do have snow. By taking the median, snowfall from MERRA or JRA55, would go into the ensemble product. On the other side of the event, MERRA might not have snow but ERA-Interim and JRA55 do, so snowfall would go into the ensemble product. This would result in a larger region receiving snow. How do you deal with that situation.

The Boisvert et al., (2018) analysis showed that the reanalyses tended to agree well in terms of the presence of a precip event, but differed strongly in the magnitude of the precip during an event. The fact the reanalyses assimilate the same sea level pressure observations means they tend to simulate storms in the same locations, , but they differ in the magnitude and phase of the precip. There was also no obvious suggestion from this analysis that there were significant biases in the location of precipitation events, although the different model grids complicate this slightly. The fact we re-grid everything onto our coarser 100 km grid should help somewhat with this issue. In general, we see this as a crude effort to produce a 'consensus' snowfall estimate on our model grid. Ideally we would have more reanalyses (that provide a direct snowfall estimate) to increase our confidence that the reanalyses are bracketing reality.

With respect to the flow and structure of the paper, sections 4 and 5 seem repetitive, especially where model sensitivity to reanalyses is discussed. Essentially, sensitivity analyses for both periods give the same results. It makes sense (to me) that if you are going to compare the two time periods then the discussion and plots are merged. This would reduce repetition, make it easier for readers to compare the two time periods, and maybe shorten the paper.

We have followed your suggestion and merged these sections, mainly through dropping most of the figures and discussion of the 1980s results. As you say, the differences are not large in terms of the model performance evaluation, and our interpretation of the different figures was pretty similar. We have kept the budget figures in the SI. We refer the reviewer to our revised manuscript which highlights this change and our justification.

Many of the figures could be improved. In many figures, the colors are not sufficiently distinct, dark purple and dark blue. This is the case with figures where dots are used. Maybe get rid of the black borders to the symbols. Also use different symbols. For many of the line graphs, increasing the weight of lines in the legend and in plots would help. Also consider whether or not you need to show the spread. The overlapping shading obscures the lines showing the means. Using shading works for two, or possibly three, series, especially if they are separated, but it starts to detract from a plot and not convey the information you want it to with more series. For example, the North Atlantic plot in Figure 11: I can't see the lower limit of the JRA55 spread or the upper limit of ASR. In many cases the spread is not discussed in the text. If you still need or want to show spread, you could just show May 1 snow depth spread as vertical bars off to the right hand side of each panel. The issue of including plots in figures but not describing the plots in the text occurs in several figures (e.g. Fig 9). I would suggest that if it not discussed, then don't include it.

We have made several changes to the figures, especially based on the comments below, including clearer legends, thicker lines, better axes. Thanks for that.

The shading appears to be an issue in Figure 11 and 13. I am keen to include the shading as it helps show where the model results may significantly differ from each other over interannual variability. While it isn't always mentioned, the presence of clear differences beyond the spread (e.g. some of the JRA and ASR regional results) is clearer than showing just the means. Showing just some shading for some of the lines would be odd in my view and questionable to the reader. We have changed the axes to prevent the cropping, lightened the shading. and changed the colours, to make this clearer. These figures should also be more colorblind friendly (dropped the green).

The spread as of May 1st is summarized in the Table, which also allows for an easier comparison across the sensitivity studies if the reader struggles to see this in the figure.

Specific comments

L3, P5. "...to avoid complexity of snow melt processes". As I note in General Comments, Fig 3 shows what could be interpreted as melt. I would like to see more justification. Furthermore, a simple temperature index approach could have been used to account for melt.

We refer to our response to the general comment above and our inclusion of more discussion about this in the revised manuscript.

Equation 1. See General Comments.

See response to that comment

Equation 3. Should this be –\nabla \dot (hu) Equation 4. Should the divergence be -h\nabla \dot u

Yes, thanks. The code was correct at least! I have changed these equations.

Equation 5. You have a wind speed threshold for wind packing but not for blowing snow. Why is this? Studies for prairie environments indicate blowing snow initiates above ~ 4 m/s, which is similar to your wind compaction threshold.

We agree it does appear odd to not apply this threshold to the blowing snow loss term, so have now added this in. This only had a small impact on the results, but has meant that we had to re-run the model and reproduce all the figures to highlight this small impact. We didn't change any of the coefficients as the changes were so negligible.

Section 2.5. Suggest this section is moved to 2.1 as an introduction to the modelling framework. This sets up the discussion of the parameterizations of the accumulation and sink terms.

We have made this change to the model description, as recommended. We hope this has improved the readability of the model formulation.

L12, P11. An advantage of reanalyses is that they produce consistent outputs. Mixing and matching fields from different reanalyses breaks this consistency. How similar are the ERA-Interim winds to MERRA and JRA55.

Yes true, but as we accumulate snow, the seasonal growth is more important than the daily variability. The study of Lindsay et al., (2014) compared several reanalyses in the Arctic (not including JRA) and showed that ERA-I winds were slightly higher (~0.5 m/s) than winds measured on drifting stations, and MERRA was slightly lower (~0.5 m/s). We decided not to add this to the investigation as it should be second order compared to the other variables and would likely just involve a recalibration of the wind packing and blowing snow loss coefficients.

L17, P11. “We linearly interpolate...” Do you mean bilinear interpolation? See General Comments. This needs more detail.

We use the Python SciPy interpolation package. The linear interpolation uses a triangular (barycentric) not bilinear interpolation approach. We have added a line: ' Gridding scripts written in Python are included in the GitHub code repository' to indicate that the reader can refer to our scripts to learn more about this and reproduce our gridding.

L13, P14. Are significant amounts of snow in summer likely to be present in recent years? The data in Warren 1999 is 30 years old at a minimum. Are there observations from N-ICE or other field campaigns to justify non-zero initial snow depths. Further- more, how do you initialize new sea ice? This needs to be explained.

The inclusion of more recent summer initial snow depths was also justified to correct a low bias in snow depths when we compared to OIB. We agree there is not much direct evidence to justify this, and the location of the N-ICE campaign will limit its utility here (highest initial snow depths north of Canada/Greenland coastlines).

The initial snow depth is only applied to regions where we have grid-cells with a concentration above 15%. As we track the snow in a given a grid-cell, new ice forms with no new snow, but can accumulate snow instantly. We have added this to the manuscript: ' New ice that forms in a grid-cell is assumed to be snow free, but these grid-cells can accumulate snow instantly.'

L24, P14. “The snow depth is distributed evenly over the old and new snow layers...”. Is there a reason why initial snow depth was not just assumed to be dense old snow.

As discussed after this line, albeit briefly, we assumed based on the sparse old observational studies that this was a combination of snow that didn't melt/persisted through the melt season, and some early summer/fresh snowfall. We admit this is somewhat unconstrained, but it based on the observations in Radionov et al. (1997)

L19, P16. “We carried out initial model calibration...” For this study, you only calibrated the model once, right? Suggest drop “initial” throughout this discussion unless multiple calibrations were made.

True, dropped initial from the text in a few places where it seemed inappropriate.

L22, P16. “...calibration involved manually tuning NESOSIM to improve the general fit...” Was this fit judged “by eye” or was some metric used? Also were all years 1980 to 1991 used, or did you leave a year out for validation during this period. While I recognize that you validated for 2000 to 2015 using OIB data, measurement accuracy and conditions might be different between the two periods.

Yeah this was by eye. We would have carried out a more optimized calibration effort if the in-situ data were more consistent, but as it was, we didn't want to over fit the model to this sparse dataset. We agree with your statement about old and new accuracies but feel this was the best compromise between producing a calibrated model we want to run primarily for a new arctic time period (because of the available altimetry missions).

L14, P17. It looks as if there is a larger difference in snow depth between January and March. Modelled snow depths gradually increase, while observed depths appear to increase in accumulation rate. Is this a shortcoming of the snowfall products. Also, you should mention the decrease in depths in April that could relate to melting and or compaction.

As discussed earlier, the limited coverage of the in-situ data prevents us from saying more about these comparisons (i.e. to be confident of the presence of a particular high or low bias in the model in our view.

L2, P18. It is difficult to believe a correlation of 0.6 for density given the spread of points in the plot.

We checked and stand by these values. The change in axes limits perhaps makes this clearer/more believable! Note that the density of points increases to the upper right which is the main reason for the moderate correlation strength.

L7, P19. “Including the blowing snow loss... but no significant change in snow density.” My first thought here, is why expect any change in density? The only mechanism by which density can be influenced by the blowing snow parameterization is a reduction in the “new snow” depth. So how deep is this “new snow” layer and how quickly does it get redistributed to the “old snow” layer?

It could change the snow density by essentially removing fresh snow (only this layer can be blown into leads) before it gets transferred into the old snow layer. We have added: ' This parameterization can impact the bulk density implicitly by reducing the amount of fresh snow contributing to the total snow depth/density.'

L7 to 15, P19. Maybe add that blowing snow loss in the central Arctic are small because sea ice concentration is close to 100%.

Adapted that line to read: 'As the drifting station data are collected primarily within the Central Arctic where ice concentrations are near to 100%,'

Section 4.2 and Figure 8. I am struggling to make sense of this section. I think part of the problem is that evolution terms are shown as cumulative, which makes comparison difficult: a big snow storm could deposit several 10's of centimeters of snow, dominating the snow depth for the rest of the season. I think you can compare the magnitudes of the terms at the end of the season (May 1) (as you do in the text) but not during the season. To compare terms during the season, I think you need to compare the timestep change in each component. The comparison is not helped by the fact that it is very difficult to distinguish lines in Figure 8. The lines in the legend need to be thicker. I would suggest leaving snow volume out of the figure.

We have tried to make the lines and shading similar, through similar efforts to Figure 11 and 13. I generally like the idea of showing the daily timestep change, although they were pretty noisy and harder to distinguish than just the cumulative plots (a lot of lines crossing over).

Our aim was really to show the cumulative impact as this fits in with the fact we validate the model in spring, at the end of the accumulation season. We agree understanding the model budget terms in all seasons is also crucial but hope to explore this more in follow up work.

L12, P24. Prefer "advected" to "drifting". For snow, drifting implies blowing snow.

Agreed, changed.

Figure 9. I would suggest showing only the evolution terms that you discuss in section 4.2. Other plots can be put in supplementary figures. While I suggest you don't show the snow volume, note that the units are a depth. Moreover, I think Ah_s (sea ice concentration * snow depth) is better thought of as a gridcell mean thickness.

In response to another reviewer we have added better labels to this figure. We are keen to keep all these panels in as we think readers could be interested in seeing the spatial importance of the various fields. We have changed volume to snow depth, and changed the final panel to 'snow depth over ice'.

L13, P26. Prefer "Soviet Station" to "old" period.

We have actually changed this to the '1980s' time period here and throughout the revised manuscript. The Soviet Station era goes back to the 1950s so didn't want to use that label to avoid confusion.

L13, P26. Given the spread in snow depths in the "New Arctic" and "Soviet Station" periods, are they really that different?

The spread in snow depths based on the input data make such conclusions hard to state and we have tried to avoid this in the paper. We have added more discussion around this issue in response to the other reviewers, but have also tried to remove the focus on comparing time differences as we think this beyond the scope of this initial model evaluation paper in line with your comments of too much regional analyses. As such we have shifted the focus to the 2000s analysis and put the 1980s results in the SI.

L8, P29. Maybe use "difference" instead of "bias" as you have no "truth".

Agreed, changed.

L2, P30. See General Comments. If Median-SF is biased it might not be that useful.

We refer to our response to the general comment.

L21, P30. “regional variability” – suggest “regional scale”. Regional scale is contrasted with Pan-Arctic Scale.

Agreed. Changed this to ' capturing the variability in snow depth at this regional scale'

Figure 17. Why does NESOSIM have zero snow depths but OIB has non-zero snow depths. It is difficult to interpret the panels with the OIB datasets overlayed. Maybe just show All-years but with separate panels for SRLD, JPL and GSFC OIB products. The individual years can be included as supplementary figures and/or discussed in the text. Also maybe use dots rather than x’s to avoid symbols overlapping.

Good spot, we realized there was a small error in the gridding and zero values in the model along the coastline were being included in the binning/regression. We have now removed these values (adding in a zero snow depth mask) which has removed this issue. This is another thing to consider in this comparison as the various OIB products have different treatments for including low snow depths, and some may use cut-offs for higher snow depths. We decided to keep all the OIB snow depths in this comparison (especially as we bin the data up to 100 km), but is something to consider in future validation efforts.

We have also taken your advice and now just show the all year regressions in separate panels. The figures of the individual years are shown in the SI and the r/rmse values are still shown in the table. We also included Kernel Density Estimate contours to highlight more clearly the differences in the distributions.

Technical Comments

Abstract, L9, P1. “Several simple parameterizations to represent key sources and sinks”. The number of processes is not large, so you might as well list them explicitly, rather than keeping the reader guessing :).

Agreed, we have added: '(accumulation, wind packing, advection/divergence, blowing snow lost to leads)' in parentheses.

L22, P3. Suggest “availability” rather than “presence”.

Agreed, changed.

L12, P4. “(Show later)” Give a figure number.

Agreed, added.

L14, P4. “Ice drift”. Suggest “Ice Motion” to avoid confusion with drifting snow.

Agreed we have changed drift to motion here and in several other places in the manuscript.

L8, P5. “...our reanalysis data...” Suggest “...reanalysis fields...”.

Agreed, changed.

Table 1. Add symbols for snow densities.

Added, along with the model variables (suggested by other reviewers).

Figure 3. As snow depths and densities are binned, could the data in the upper panels be shown as “box and whiskers” or just boxes. That way readers can see the amount of overlap between depth and density estimates. I suggest you spell out Soviet Stations in the figures. Add 1:1 lines on the lower panels. It would be nice to have a single symbol in the legend.

We have added Soviet Station to the legend and axis labels, reduced number of markers to 1, and added the 1:1 line. We decided box and whisker plots were too much as we already show the mean and spread (+/- 1 SD) and have the raw values below.

Figure 4. Does No Initial Conditions (NO IC) mean the model was initialized with 0 cm snow depth?

Yes! We have added more information to the caption to help clarify the model runs.

L1, P26. Shouldn’t this be Section 5?

Yes! changed.

L13-14, P29. Reference needed.

Added reference to Boisvert et al., (2018).

Figure 16. Why two symbols in the legend. Also the colors are difficult to distinguish. Maybe no black border on symbols. Also use different symbols.

We have changed this to just show one marker, made the markers bigger and made the OIB markers squares to more clearly differentiate them from the model markers.

Author response to Anonymous Reviewer #2 on: "The NASA Eulerian Snow on Sea Ice Model (NESOSIM): Initial model development and analysis" ***by*** **Alek A. Petty et al.**

Reviewer comments are in black, our responses are in blue.

We will also submit the revised manuscript and a word document highlighting the tracked changes we have made based on these comments.

This clearly written manuscripts provides a detailed description and exploration of the new NESOSIM snow model. NESOSIM produces gridded, daily snow thicknesses and densities for Arctic ocean sea ice during the accumulation season (defined as mid- August through April) given daily inputs of Arctic wide snowfall, sea ice concentration, ice drift, and near surface winds. Although the Arctic melt season may extend well into September, the model does not include thermodynamic or radiative processes, and this certainly limits its utility. Rather, the emphasis here is on the impacts of wind via wind packing and blowing snow loss to leads/open water. The parameterizations are fairly simple – winds exceeding a threshold can only decrease snow thickness and increase snow density. There are no snow drifts, for example, or sub-grid regions of bare ice which are present in other models. In addition, there are no snow-aging processes that may contribute to density changes. Still, the authors do a commendable job validating their model against observations and do a thorough evaluation of model sensitivity to the various snowfall reanalysis, ice drift and ice concentration products. This latter analysis highlights the true utility of the model – a simple framework for the inter-comparison of reanalysis-derived snow on sea ice data products.

We thank the reviewer for their time in producing this review and the useful comments provided.

Some specific scientific comments: The authors need to better place the work in scientific context and show how the work is unique. How is this an improvement over the simple models of snow depth forced from reanalyses? There are more complex snow on sea ice models (Lecomte-LIM, Liston-SnowModel, Hunke-CICE) which include some of the same processes (ice drift, dynamics, precipitation) yet rather than develop wind loss and compaction include some distinctly different processes (thermodynamics, radiative properties, snow ice formation, dune formation, ridge accumulation. . .). Are these models missing the "key sources and sinks"? There is also Dery and Tremblay (2004, JPO) that specifically looks at the effects of wind redistribution with an explicit mass flux into leads. Is your approach better? More useful? Consistent?

This is a fair comment, although we want to stress we don't seek in this paper to produce a model that is 'better' or 'more sophisticated' than the models currently available. Indeed this would be a real challenge considering the complexity of some models already available. But model sophistication isn't the only factor in producing reliable snow depths - the forcing is arguably more important, and there is still a high level of uncertainty surrounding the sensitivity of snow to various forcing data, especially snowfall. As such we sought o develop a model that will enable us to explore these sensitivities and increase model sophistication as needed based on these uncertainties.

We have moved up to the introduction and added more text regarding our motivation/philosophy (this was at the start of the modelling description) to make this clear from the outset and have also adapted the introduction to add the following before this statement:

" Due to these observational limitations, the sea ice community often relies on simple models of snow depth forced by reanalyses (primarily snowfall data) (e.g., Maksym and Markus 2008;

Author response to Anonymous Reviewer #2 on: "The NASA Eulerian Snow on Sea Ice Model (NESOSIM): Initial model development and analysis" ***by*** **Alek A. Petty et al.**

Reviewer comments are in black, our responses are in blue.

We will also submit the revised manuscript and a word document highlighting the tracked changes we have made based on these comments.

This clearly written manuscripts provides a detailed description and exploration of the new NESOSIM snow model. NESOSIM produces gridded, daily snow thicknesses and densities for Arctic ocean sea ice during the accumulation season (defined as mid- August through April) given daily inputs of Arctic wide snowfall, sea ice concentration, ice drift, and near surface winds. Although the Arctic melt season may extend well into September, the model does not include thermodynamic or radiative processes, and this certainly limits its utility. Rather, the emphasis here is on the impacts of wind via wind packing and blowing snow loss to leads/open water. The parameterizations are fairly simple – winds exceeding a threshold can only decrease snow thickness and increase snow density. There are no snow drifts, for example, or sub-grid regions of bare ice which are present in other models. In addition, there are no snow-aging processes that may contribute to density changes. Still, the authors do a commendable job validating their model against observations and do a thorough evaluation of model sensitivity to the various snowfall reanalysis, ice drift and ice concentration products. This latter analysis highlights the true utility of the model – a simple framework for the inter-comparison of reanalysis-derived snow on sea ice data products.

We thank the reviewer for their time in producing this review and the useful comments provided.

Some specific scientific comments: The authors need to better place the work in scientific context and show how the work is unique. How is this an improvement over the simple models of snow depth forced from reanalyses? There are more complex snow on sea ice models (Lecomte-LIM, Liston-SnowModel, Hunke-CICE) which include some of the same processes (ice drift, dynamics, precipitation) yet rather than develop wind loss and compaction include some distinctly different processes (thermodynamics, radiative properties, snow ice formation, dune formation, ridge accumulation. . .). Are these models missing the "key sources and sinks"? There is also Dery and Tremblay (2004, JPO) that specifically looks at the effects of wind redistribution with an explicit mass flux into leads. Is your approach better? More useful? Consistent?

This is a fair comment, although we want to stress we don't seek in this paper to produce a model that is 'better' or 'more sophisticated' than the models currently available. Indeed this would be a real challenge considering the complexity of some models already available. But model sophistication isn't the only factor in producing reliable snow depths - the forcing is arguably more important, and there is still a high level of uncertainty surrounding the sensitivity of snow to various forcing data, especially snowfall. As such we sought o develop a model that will enable us to explore these sensitivities and increase model sophistication as needed based on these uncertainties.

We have moved up to the introduction and added more text regarding our motivation/philosophy (this was at the start of the modelling description) to make this clear from the outset and have also adapted the introduction to add the following before this statement:

" Due to these observational limitations, the sea ice community often relies on simple models of snow depth forced by reanalyses (primarily snowfall data) (e.g., Maksym and Markus 2008;

Kwok and Cunningham, 2008; Blanchard-Wrigglesworth et al., 2018). More sophisticated snow on sea ice models are available, such as SnowModel, a terrestrial snow model recently adapted for sea ice environments (Liston et al., 2018), as well as the snow parameterizations included in sea ice climate model components, such as CICE (Hunke & Lipscomb, 2010) and the Louvain-la-Neuve Sea Ice Model (LIM), which has recently undergone various improvements to its snow physics (Lecomte et al., 2015)."

What is the impact of excluding thermodynamic processes on your results? Does this change your conclusions about the impact of wind processes?

Unfortunately we really lack the data needed to better answer this question. It's likely that including thermodynamic processes (e.g. snow melt) will reduce the impact of the wind loss term, but that is already a very unconstrained process in this model.

We also refer the reviewer to our response on this subject of missing melt processes raised by Reviewer 1 and the newly added discussion of not including snow thermodynamics, the potential impact of this, and the future work section at the end of the paper.

Some misleading statements: First sentence of the abstract. . . . "produces daily esti- mates of depth and density of snow across the polar oceans". Not yet because of some important missing processes. Qualify with Arctic only and during the accumulation sea- son.

Agreed. Changed 'polar oceans' to 'Arctic Ocean through the accumulation season'.

Using old vs new snow in the text and figures. It's clear that there is intention to one day include snow aging, but for now there is only fresh vs compaction. The depth hoar densities of 150-250 is never used in the model even though paragraph 10 seems to suggest that it is. The old snow value is 350 kg/m3 which is not the average of the higher end of wind slab and depth hoar (325 kg/m3) but rather the average of the wind slab bounds.

The reviewer is correct that there is the intention to one day include snow aging in the model, but, for now, there is only "new" and "old" snow. We think there may be some confusion here as we take a weighted average based on the average ratio of depth hoar and wind slab from Radionov et al. (1997) and Sturm et al. (2002). The density values chosen for these layers are based on the literature on snow density observations (e.g., Radionov et al., 1997; Sturm et al., 1998; Sturm et al., 2002; Sturm and Massom, 2017) and implicitly include values for density layers that are not explicitly treated in the model. For example, the compacted layer uses a value of 350 kg/m3, which is a value calculated using the average ratios of depth hoar (40%, 250 kg/m3) and wind slab (60%, we found values of 410 kg/m3 in the referenced works above) within a snowpack. Admittedly, the ratio of depth hoar and wind slab seasonally changes and the model does not account for this seasonal change. However, we feel this is a better treatment than neglecting the influence of depth hoar altogether in the "old" snow layer. We have edited this description in the revised manuscript to make this clearer.

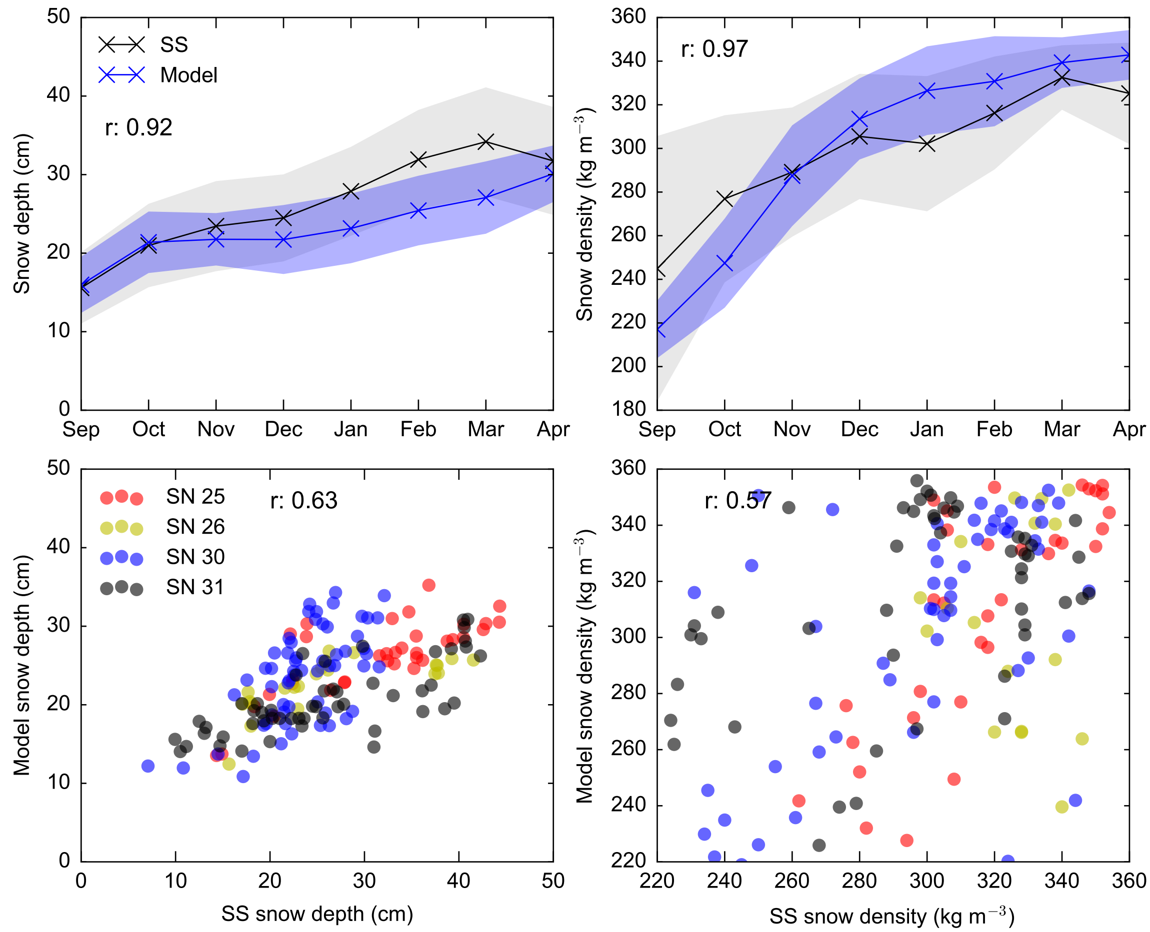
Perhaps future developments could be kept to a specific section to better clarify what the model does and doesn't do.

Yes, Reviewer 1 wanted this too, so we have added this to the final summary section. We refer the reviewer to this new subsection.

Snow density in NEOSIM is bounded by the two chosen snow density parameters (200 and 350 kg /m3) even though the observations referenced give values for dry snow of 150 and wind slab ~400 kg /m3 on average. Why exclude these possibilities at the outset? Instead of using an average value, doesn't it make more sense to use the upper and lower bounds given the nature of the parameterization? How sensitive is the model to these values?

We did experiment with these values, but found that this resulted in an over-estimation of the strength of the snow density cycle. See figure below. We have added this to the new manuscript:

"We did experiment with alternative snow densities (e.g. the wider spread of 150 and 400 kg m^{-3}) but found this provided worse correspondence with the seasonal snow density evolution compiled from in-situ Soviet Station data (introduced in Section 3.4)."



The late summer initial conditions integrate all the missing snow melt processes and for that reason, they are rather important. The paragraph on page 14 does a fairly good job motivating your approach, but it would be clearer if you showed the equations for hs(0) and hs(1) after summer melt. Also better explain how snowfall events factor into this parameterization and explain why keeping the same fraction for fresh/compacted snow is the right approach (or clarify if you do something different). It would also be informative to see the Aug 15 values in your figures 3 and 4. Are there Aug observations to help validate the IC parameterization and fig 2 in particular?

We think there may be some confusion here. We rerun the model every August, regardless of the prior spring snow conditions. The starting point for the model is a direct application of the initial condition snow depths.

We don't show these values in the figure because we start the model halfway through August and these figures are showing monthly means/spreads.

The equal split between fresh/old snow was due to (albeit fairly crude) reasoning that some of this initial snow may be from snow persisting through the melt season and some due to new snowfall through summer based on observations in Radionov et al. (1997). We explain and cite this in the manuscript.

We lack consistent observations of August snow depths to validate this component of the model, but we did look at the snow depth data from ice mass balance buoys – some show the presence of snow in August, while others do not, within the same summer seasons. While these observations suggest snow is present in August, these are point measurements and may not be wholly representative of basin-scale snow depth distributions.

Why absorb the timestep in the model equations? In (2) the parameter alpha has a timestep dependence that isn't explicitly called out and as a result, 0.05 is less meaningful. Better to define an alpha with units of per second.

Agreed, we have changed these equations to include a time step value, so these coefficients are now independent of the length of the time step. The units have been updated accordingly (Beta needs to be in units of per meter as we multiply by wind speed).

Are the differences between simulations with different snowfall estimates larger than the differences between time periods? Are the time period differences significant?

The spread between reanalyses makes it challenging to ascertain confidence in a physical system change between periods. Understanding the factors contributing to the differences between periods and reanalyses (e.g. the contribution from precip, freeze-up, other processes) is beyond the scope of this paper, but is a topic of ongoing work.

The potential changes in snow depth and their physical cause are part of a hopeful follow-up study which can elaborate on this in the detail required. We also refer you to our response to Reviewer 3 where we provide a brief summary of our thoughts about this.

Note also that to shorten the analysis description (and to prevent the focus being about changes in snow depth) we have merged the 1980s/2000s reanalysis sensitivity study sections and moved the 1980s figures of model evaluation to the Supplementary Information. The 1980s reanalysis sensitivity studies did not add much to our discussion and interpretation of the model performance so we have dropped this from the revised manuscript.

Fig 14 seems to suggest that ice drift is actually quite important but masked by basin or large regional averaging. Magnitudes of the differences are similar to the snowfall sensitivity. Impacts are near the ice edge (increase ice retreat?) and add to smaller (but still > 100 km) scale variability (potentially impacting melt-pond formation).

Agreed! We discuss this in the paper already so haven't added more here.

Technical corrections:

Table1. add the model variable in the table.

We have added all model variables to the table.

Define U in Eq. (5).

It was defined after Eq. 2 earlier.

Missing) in Eq. (6)

Added

Why is the bs term in Eq. (7) positive?

The wind loss term is now negative, so this should make more sense now!

Missing t dependence in some terms in Eq. (7) and the next equation Line 26 missing "of" in "one the better.."

Thanks, good spot. Added.

Add W99 to upper panels of fig3 Add W99 to fig 6 (a)

We decided not to add W99 to figure 3 as the legend makes clear we show data from the drifting Soviet Stations (hence SS in the legend) from 1981 to 1991. The W99 is a climatology of these data over the longer time period (back to the 1950s) so it is not exactly the same. Figure 6 shows no Warren or Soviet Station data so we didn't add this either.

In explanation for fig 8, NA region shows a small "decrease" due to convergence, not increase. CA shows a small increase but is not mentioned.

Changed this to: 'The NA region also shows a small (~2 cm) decrease (increase) in snow depth driven by snow/ice divergence (ice/snow advection), while the CA region shows a small (~2 cm) increase in snow depth driven by snow/ice convergence.'

Fig (9). What is (b) Ocean? Change (d) wind/leads to wind loss to leads. Is there an ice area cutoff for snow depth in (k)?

Ocean was the snowfall into the ocean. We have updated all these labels to make this a lot clearer, including the variable names and a link in the caption to the table which provides the model variables.

All snow depth results and budget fields shown in the analysis use a concentration cut-off of 15% and which we have now added to the revised manuscript at the start of page 9 after the model formulation description: 'Note that this bulk snow density is masked if the respective ice concentration is less than 15% or snow volume is less than 2 cm, while all snow budget terms presented here show data only when the concentration is above 15%, to prevent spurious results in regions of near open water conditions.'

Table 4 could be improved by adding the 1981-1991 time period for comparison, identifying the boxed regions as “reanalysis sensitivity”, “ice drift sensitivity”, “ice concentration sensitivity” and including in the description that the default configuration is MEDIAN-SF, NSIDCv3 ice drift and Bootstrap ice concentration

See comment above. The comparison between time periods is the focus of on-going work led by co-author of this paper Melinda Webster. As it requires careful evaluation of the uncertainty in forcing data and changes in these fields we feel we can't make a simple comment about that here. We have thus decided to drop the 1980s reanalysis figure (the 2000s figure serves this purpose just as well) and moved the 1980s budget figures to the SI.

End of p 30. Comment on the NA region in table 4 when ice drift is included. Seems to be important here too.

Yes, this was an omission and has been added to the text.

Fig 13, (d) NA does not appear to be consistent with table 4. The NODRIFT value on May 1 is around 34 cm in table 4 but seems to be much higher in this figure. Explain.

The table value was wrong and has been corrected. Although based on the tweak to the wind loss term (introduction of the wind action threshold for this term as well as wind packing based on reviewer recommendation), all values have been reproduced and are slightly different from their original values. Apologies!

Last comment before 5.3 does not seem correct. There does appear to be bias in the real-time products with respect to the peripheral seas .

Agreed, we have dropped this statement.

P 37. What are uncertainties in the OIB observations?

The recent STOSIWIG study (Kwok et al., 2017) stated that snow depth can be estimated from a Snow Radar echogram with an uncertainty of 'several centimeters' although this depends strongly on the ice conditions, the particular Snow Radar system being used, and various other factors (e.g. geolocation errors associated with the plane pitch and roll. We have added a sentence regarding this to the discussion at the end of Section 3, where the OIB data are introduced.

Mention the results from sensitivity of ice concentration in the summary. These were interesting and significant.

OK we added the following: 'We also briefly assessed the sensitivity of NESOSIM to the input concentration data, with our results suggesting that this choice of product (Bootstrap and NASA Team explored in this study) can have a significant impact and should not be overlooked.'

Author response to Anonymous Reviewer #3 on: "The NASA Eulerian Snow on Sea Ice Model (NESOSIM): Initial model development and analysis" ***by*** **Alek A. Petty et al.**

Reviewer comments are in black, our responses are in blue.

We will also submit the revised manuscript and a word document highlighting the tracked changes we have made based on these comments.

The authors present a new open source model, the NASA Eulerian Snow on Sea Ice Model, for estimating daily depth and density of snow on sea ice. The authors note at a few points in the paper that the model is being developed primarily with application to altimetry-based ice thickness determination in mind, though other applications are likely. The model is a simple representation of the snow that is largely an accounting of snowfall produced by reanalysis data, similar to prior efforts (e.g. Maksym and Markus 2008; Kwok and Cunningham, 2008), with terms for snow compaction, loss to leads, and transport on sea ice. It is Eulerian, but features pseudo transport by exchange between grid cells, features only 2 layers, and is forced with available spatially and temporally complete datasets that are known to be of limited accuracy (e.g. Reanalysis, passive microwave concentration). The model is calibrated/validated against limited available snow on sea ice data from Operation Ice Bridge and from 1980s era Soviet drifting stations. The description of the model is complete and in this regard the model is publishable with minor revisions – but reviewer doesn't feel the model is very good or useful in its current form for its intended purpose. Reviewer focuses most of this review on highlighting its shortcomings. In fact, a possible conclusion of this data presented would be that simple treatment of snow on sea ice will not meet the accuracy levels required for altimetry applications. The reviewer encourages the early career team to put the paper aside for awhile and take the time write a model that would actually be highly used.

We thank the reviewer for putting the time into providing this review. As also highlighted by the other reviewers, the model we developed has provided a useful means of thoroughly assessing reanalysis-derived snow depths and their sensitivity to the input forcing data used. It is hard to develop a highly sophisticated model and explore true sensitivity to input forcing data, and the latter has been extremely lacking in the literature to-date. We believe this study has provided the needed baseline from which we hope to increase model sophistication in future as needed.

The reviewer feels that the key issues are that the model is excessively simplistic, not representative of known physical process (even at the level of simplicity targeted), and that its results show it is inadequate for the intended purpose. There are errors in the equations presented, many compromises appear to have been made that make accuracy and/or realism lower in favor of rapid release, and as a result the work is unlikely to have much impact as presented. The presentation in the paper is quite long, and focuses on trying to convince the reader that the model is good, rather than taking a hard look and comparing against a reasonable standard.

We know of no other pan-Arctic snow depth & density product that provides significantly higher accuracies (e.g. the rmse when compared against OIB say).

Based on the comments of Reviewer 1 we have added similar comparisons of the OIB snow depths against Warren and modified Warren climatologies, highlighting similar/better agreement when using NESOSIM (in terms of correlation coefficient and/or root mean squared errors). We are somewhat limitied by the uncertaintiy in the OIB products (several centimeters) that prevent us from carrying out a more complete validation.

We have also added a lot more justification at the start and end of the revised manuscript regarding our approach and expectations for future model development/calibration efforts to improve these error.s

We have also added the following to the end of the abstract: ' Potential improvements to this initial NESOSIM formulation are discussed in the hopes of improving the accuracy and reliability in these simulated snow depth and density'

The development of a snow product for improving retrieval of sea ice thickness from altimetry is critical for ICESAT 2 to be useful and this team should have NASA's support to do just that. Such a snow model's accuracy goal must be based on a desired accuracy in thickness retrievals. (e.g. retrieval of ice thickness accurate to +-0.5m over a given domain demands snow depth accurate to O5cm over the same domain). The model presented here is not up to meeting these kinds of needs, and does not leverage the existing (more sophisticated) models of snow on sea ice (e.g. LIM, SnowModel, CICE).

We are not aware of more sophisticated approaches providing the requirements you list, and our belief is that uncertainties in the input forcing data needed to be explored further, along with efforts to improve model sophistication. We see this model as a contribution towards this end goal. Based on the comments of the other reviewers we have added more discussion of the other snow models available and the physics they include.

We agree that working towards a 5 cm error makes sense, and we believe this provides a useful framework to build towards that goal. This is also discussed in a new study by a co-author of this study: Webster et al. (accepted) which speaks to several issues of our current abilities in observing and treating snow in models. Existing basin-scale observations unfortunately do not have 5-cm accuracy; snow depths derived from merged satellite data do not have 5-cm accuracy. So it is difficult to produce model output with 5-cm accuracy given that no observations of 5-cm accuracy (or better) at the basin-scale exist for model validation and assessment, and especially on a seasonal time-scale.

Some major issues include: Model design relative to state of knowledge: 1. The two layers used in the two layer model (new snow and windslab) do not represent the two layers of the snowpack discussed in literature (wind slab and depth hoar). Authors cite and discuss the literature indicating that windslab and depth hoar dominate the mass of the pack and have quite different density – then ignore these decades of observation to invent a new scheme unsupported by observations. Respecting the effort to create a simple, 2 layer model, new snow should not be one of the two layers. The references cited clearly state that new snow rarely comprises much of the Arctic snowpack, because it is very rapidly converted to windslab. The preservation of a new snow layer appeared to be designed for modeling loss of snow into leads – but little is known about the magnitude of this flux, and it was minor in this model. 2. The model is operated on a 100x100km grid, which is very coarse relative to the variability in ice – which is shown to be important in impacting the accumulation of snow. The data sets used provide much higher ice concentration, and movement information – this data should be used at full resolution and atmospheric data can be downsampled. 3. Melt is neglected despite it being important during part of the timeframe and having significant impact on results.

1. We expanded on the description in the manuscript about the second layer representing depth hoar and wind slab and their respective densities. We treat wind slab and depth hoar by taking the average ratio of these two layers, based on Radionov et al. 1997 and Sturm et al. 2002, over the

accumulation season. This ratio is then used to take a weighted average of the wind slab and depth hoar densities. We agree that new snow provides a smaller fraction of the total snow, especially towards the end of the accumulation season, which occurs in both the model and in observations (Warren et al., 1999).
2. We respectfully disagree, especially as we are taking gradients in ice drift fields on daily time-scales that can be very noisy, so other recent studies, e.g. Holland et al., (2014) smooth the data before using them. We hope to explore specific high res configurations of this model in the future, but wanted to carry out more extensive sensitivity studies across various forcing data in this initial model development stage.
3. Yes we neglect it and have added more discussion on this point based on the input from reviewers 1 and 2 in the model development and future priority sections.

Quality of the Model Results and Characterization thereof 1. Validation shown indicates the model produces results that do not capture the variability in observed snow depth or density reliably. Authors focus on averages of model output over decadal timeframes, which can be made to match observations by tuning of the arbitrary, non- physical constants in the model. This focus fails to acknowledge the inability of the model to capture interannual or spatial variability. 2. Prediction intervals are not pro- vided, but scatter plots show little relationship between individual observations of snow depth and modeled snow depth. No discussion is provided of how these errors would propagate in the intended use (altimetry retrievals of ice thickness) but it appears errors are sufficient to radically alter retrievals of depth and appear to indicate the data would not be useful for altimetry retrievals of ice thickness from ICESAT2. Authors fail to acknowledge any of these shortcomings and go to great pains to make the results appear good. 3. Modeled variability in density appears to have very little relationship to observations. 4. Comparison with the southern ocean, are pushed to a future effort, but validation statements in the paper suggest the model applies to 'polar oceans'. 5. Results from the median of the three reanalysis products are declared 'better' repeatedly with no reasonable support. Taking the median of atmospheric reanalysis models would result in nonphysical jumps between atmospheric states and the removal of extreme events from the record, and is challenging to support physically.

1. We are confused by this comment as the assessment of how well the model captures spatial and interannual variability is shown in the OIB validation section.
2. Not sure what you mean here either. We show both the correlation coefficient (how well it agrees on the relationship between the two distributions) and the root mean squared error which provides a model error compared to OIB. We did not hide these numbers and have been very open with the performance of the model. The RMSE of ~9-10 cm is clearly not ideal but a value like this was broadly expected considering the challenges inherent in modelling snow accumulation and snow depth/density. The observational uncertainty on the order of several centimeters should also be considered here and we have made note of this in the discussion. It's thus hard to translate this into a snow depth uncertainty.
3. Correct, as we stated in the manuscript.
4. We have made clearer this is focused on the Arctic Ocean.
5. We have made clearer the justification for showing more of the median results in the New Arctic time period discussion was mainly for simplicity!

DETAILED COMMENTS

Page 1, line 16. "very strong agreement" Delete "very strong"

It does show very strong agreement with the seasonal cycles (very high correlations with the seasonal correlation plots) which is what we are referring to here.

Page 1 line 22 descriptions of agreement too subjective. The use here is altimetry. Tell the reader about the error in estimates implied.

We have changed this to: showing moderate/strong correlations and root mean squared errors of ~10 cm depending on the OIB snow depth product analyzed. These are similar to the comparisons of OIB-derived snow depths and the commonly used modified Warren snow depth climatology. Potential improvements to this initial NESOSIM formulation are discussed in the hopes of improving the accuracy and reliability of these simulated snow depths and densities.' We are wary of translating this into an error, due to the uncertainty and differences in the OIB snow depth products.

Page 2 line 5-8. Poorly worded sentence. Consider modifying. One suggestion is: The altimetry technique involves measurements of freeboard, the extension of sea ice or snow surface above a local sea level. Estimates of snow depth are required to derive sea ice thickness from either snow surface freeboard or ice freeboard, because snow depresses ice freeboard and adds to snow surface freeboard. Snow depth is one of the primary sources of uncertainty for both laser and radar altimetry (e.g. Giles et al., 2007).

We appreciate the suggestion but prefer the sentence as is.

Page 2 line 10. Replace 'lacking' with something more descriptive/accurate (they aren't lacking they are just not complete/good enough).

Changed this to 'very limited'. Also added 'direct' to observations at the start of the line.

Page 2 line 22-24. The sea ice community often relies on simple models of snow depth forced by reanalyses – please clarify how this is different. To the reader, it still looks like a simple model forced by reanalyses!

Here we are just making a comment here about what the community are using currently. Ours is also a simple model forced by reanalysis as you say and we don't think this suggests otherwise.

P 3 Line 16 "and two snow layers to broadly represent the evolution of both old/compacted snow and new/fresh snow." The assignment of the two layers in this two layer model is not consistent with the widespread understanding of the primary two layers on sea ice as depth hoar and windslab. New snow is occasionally present but usually rapidly transformed to windslab. It may be an acceptable third layer. See many of the snow on sea ice references cited here, such as Sturm et al., 2002 – generally the snow is treated in these two layers. The author's choice here to take the two layers to represent layers that the extensive literature reviewed does not discuss is perplexing.

We completely agree with the reviewer that the primary snow density layers are wind slab, depth hoar, and to a much smaller degree, new snow which, in reality, is more like a "transient" snow layer that gets redistributed by the wind. If reduced to two layers, the snowpack would be wind slab and depth hoar.

Given that the model does not have a temperature dependency, we were not able to parameterize a depth hoar layer since this is dependent on the temperature gradient within the snowpack.

Instead, we chose to include a “new snow” second layer, representing recent snowfall and blowing snow, which reduces the bulk density of the snowpack. Likewise, for the “old” snow layer, we explicitly chose density values that result from the mean ratio between wind slab and depth hoar from works by Matthew Sturm and historical data from Radionov. Although the model doesn’t explicitly take into account the seasonal cycle of this ratio, we feel that it’s a better treatment than applying the higher-end density value of wind slab and ignoring depth hoar altogether. Related, we chose not to apply a bulk climatological density because of its questionability in representing regions where observations are lacking.

P2 line 18 replace “detailed” with “iterative”. The simplified scheme does not permit a ‘detailed’ assessment of connection between input data and snow depth given its lack of physical complexity – it permits an easier iteration of possibilities.

Removed detailed from this line.

P4 line 13 Input data from passive microwave higher resolution than 100x100km, even if atmospheric data is not. Since ice concentration is so important, reviewer questions if 100km resolution is adequate. Further - does observed snow depth vary over 100km resolution? Since this is the motivation, what resolution is needed for useful for altimetry based determination of sea ice thickness?

As both drift and snowfall products come on grids with resolutions above 60 km, we did not want to use the higher res 25 km ice concentration grid. We think in these budget models the grids should be at least as coarse as the coarsest input data or it could look misleading. This is the approach taken in the concentration budget studies referenced in the manuscript. In reference to the later part of the question, higher resolutions are clearly better, but accuracy is really more important. One can always downscale data to higher resolutions but doing so doesn’t add additional information at that higher resolution.

Page 4 line 14 add “from reanalysis data” after the word ‘drift’.

The ice drift is not from reanalysis data. We have modified this to 'The model is forced with daily data of snowfall and near-surface winds from reanalysis data, satellite passive microwave ice concentration, and satellite-derived ice drifts.'

Page 4 line 16 – (volume of snow per unit grid cell in units of meters) – doesn’t make sense volume is meters cubed. Throughout the treatment of snow varies between depth and volume freely, but this free transition between volume and depth is challenged for some considerations of snow – particularly convergence/divergences. Since the goal here is to understand depth for altimetry retrieval, a convergence, which moves volume into a cell, is not the same as a change in depth.

This kind of terminology is common in models (e.g. CICE) to describe a quantity that is expressed over the entire grid-cell. Based on the recommendation of reviewer 1 we have tweaked the terminology used so that we make it clear we track the effective snow depth (over the entire grid cell) but can derive the snow depth over the ice fraction by dividing by ice area.

Page 5 table one – put formal references to data sources, e.g. “bootstrap” is not sufficient.

We provide references in the text and in the data section of the paper.

Page 5 delete “snow pit and density data. . . helped guide. . . parameterization . . . seasonal evolution.” There is no prescribed seasonal evolution of density, use of snow pit data etc. in this model. Two constant snow densities are selected and declared. This sentence obfuscates the very simple, non-experimentally supported nature of the scheme.

The model produces a prognostic density from the ratio of old and new snow and the calibration of the model was guided by this data. The average ratio of wind slab and depth hoar within the second (old snow) layer is parameterized based on snow pit observations from Radionov et al. 1997 and Sturm et al. 2002.

Page 6 line 8 replace bulk density with mass.

Agreed, changed.

Page 6 – here authors note that the community of snow science experts and prior literature they have created generally group the snow into two layers (wind slab and depth hoar). They further note substantial differences observed in density of these two layers, and that these two layers comprise the majority of the snowpack. Not noted, but available in the literature is data showing that the contribution of the two layers to the overall snowpack varies from the approximately 50-50% contribution seen at SHEBA. So it seems windslab and depth hoar are the two layers to model. But. . . these two layers are different than the layers the authors have chosen (new snow/old snow). It seems a major departure from decades of snow research is being made here and it is not being well defended. Why?

We discuss this in more detail in response to some of your earlier comments, but briefly we were aiming to keep the model simple in this fist iteration. As we have no snow thermodynamics, explicitly capturing both snow layers and their different densities is less important for our given purpose.

Page 6 line 12 “for this reason we use the average of higher end values of ws and dh”. Reviewer sees no reason provided supporting the use of the higher end of the range of values for each of the two common layers. The mean density of each layer, multiplied by the mean fraction of each layer should provide a more representative density for the combined wind slab and depth hoar. Further, the value selected is not the average of the higher end of the range of values for each of the two common layers, leaving it unclear how it was determined.

The density values chosen for the wind slab and depth hoar in the “old snow” layer are based on the average ratio of these properties within the snowpack from Radionov et al., 1997 and Sturm et al., 2002. The compacted layer uses a value of 350 kg/m3, which is a value calculated using the average ratios of depth hoar (40%, 250 kg/m3) and wind slab (60%, we found values of 410 kg/m3 in the referenced works above) within a snowpack. Admittedly, the ratio of depth hoar and wind slab seasonally changes and the model does not account for this seasonal change. However, we feel this is a better treatment than neglecting the influence of depth hoar altogether in the “old” snow layer. We have revised this description in the manuscript to make this clearer. In the conclusion, we expand the discussion to include future work in incorporating a temperature-dependency of the model, which will enable the separation of wind slab and depth hoar layers rather than applying a crude treatment for the old snow layer.

Page 6 Line 16 “Our simple parameterization is thus expected to be generally representative” No reasonable evidence provided supports this. Statements like this are found throughout this paper.

Delete or support with concrete evidence that quantifies what the range of uncertainty they will work within.

We have deleted this statement from the revised manuscript.

Page 6, Line 23 (default of 5m/s). Default or for the purposes of this work is it simply always set to this?

It's fixed in the model but can easily be changed in the open source code.

Page 6, Line 24 “determines the fraction... transferred...” Over what time? (seems that the coefficient is model timestep dependent. . . and perhaps shouldn’t be)

Yes, we have changed this to not be timestep dependent (units of s^{-1} now)

Page 6 line 26 ‘Wind threshold of 5m/s was determined based on. . .’ studies. Please add a description of the range of wind thresholds indicated by these studies, and why 5m/s was selected from within that range.

We implemented a 5 m/s threshold based on Liston and Sturm (2002) and the dry snow transport in Li and Pomeroy (1997). In reality, this threshold depends on the snowpack’s physical properties (grain size, water content, etc.) and atmospheric conditions (humidity, air temperature, etc.).

We conducted sensitivity tests on the wind speed threshold and found this to have a negligible effect on the modeled snow depth distributions. However, there are some regions where the wind threshold and wind loss term will play more important roles, such as the Antarctic environment where more leads are present, more snowfall events occur, and windier conditions occur more frequently relative to the Arctic (Massom et al., 2001; Toyota et al., 2016; Massom and Sturm, 2017; Massom, pers. comm., 2018; Webster et al., accepted).

Page 6 Line 8 Daily gridded ice drift is still required in this Eulerian scheme, eliminating it as a reason for choosing Eulerian over lagrangian, discussed above.

Our statement referred to the fact you do not need consistent ice drift - i.e. the model can be run for periods/regions with no ice drift.

Page 7, line 19. Reviewer is not aware of any evidence indicating that the loss of snow to leads in the North Atlantic sector of the Arctic is significant relative to the thick snowpack in that region. No evidence seems to be coming out of the N-ICE experiment to that effect. Some quantification of loss to leads in the Antarctic has been made by Leonard and Maksym as noted, but this was in the southern ocean. Please cite appropriate literature or delete speculation.

Correct, the snow lost to leads in the North Atlantic may not be significant relative to the thick snowpack in that region. However, we speculate that a greater proportion of snow is lost to leads there than in other Arctic regions given the lower ice concentrations, more open leads, more frequent snowfall events making more fresh snow available to redistribute, and windier conditions in the North Atlantic.

Page 8 line 4 – This parameterization doesn’t make sense and is under supported for several reasons. 1. It appears that a constant coefficient beta is multiplied by 10m windspeed NOT by the amount which the wind speed exceeds the threshold velocity! So snow is lost to leads even when winds are too slow to move snow. 2. The amount of the snow lost to leads increases linearly with windspeed, when the drifting snow volume is well known to vary more rapidly than linearly 3. The loss to leads varies linearly with open water area, again this is likely more rapid than linear, and a thought experiment with random lead spacing/size could arrive at a better approximation. 4. The parameterization removes a fraction (2.5%) of the new snow layer to leads on each windy timestep – timestep is then important due to compounding what timestep is this defined for? 5. Is this parameterization/ value supported by any field quantification of loss to leads or is it simply made up due to lack of available observation. Either is fine, but state which it is. Page 6 line 9 – missing parenthesis on equation

1. We have changed this such that the wind action threshold is applied to the blowing snow loss term.
2. Our approach is simple, agreed. we hope to explore each term in more detail in future developments.
3. Agreed.
4. This term is now time step independent.
5. Yes, this was not based on any studies. We have added clarification to the text: ' We have no observational constraints for this parameter and is a free-parameter in this model, chosen through our model calibration efforts.'
6. Added the parenthesis, thanks.

Page 6 Equation 7 – appears incorrect. Change due to blowing snow is added (last term), but this should be a loss term (loss into leads). It appears that the term calculated in Eq 5 is always positive, so adding here will result in addition of snow, not loss. Similarly, how signs are handled on dynamics, convergence and divergence as well as advection depends on how (+-) ui is defined in equation 3 and 4, and this is not (but should be) specified above. . . so the reviewer is unsure if the sign here is handled correctly.

Correct, this was a mistake. We have added a negative sign to Eq 5 and 6. This was correct in the model code. Dynamics are all correctly signed.

Page 8 line 21 August is mid- late summer. Change "early" to 'late' or delete.

Changed this to 'middle'.

Page 9 line 2-3 Do these melt events invalidate the results here? Is this model useful before these 'hoped for' additions occur? It sounds like this is being hurried along.

We have no data to test if this invalidates the results here but the extra snowfall did improve our initial model testing efforts so we included it despite the obvious concerns of missing melt events.

Page 9 line 8-13 This paragraph appears to handle a specific test case, not discussed here. Seems out of place possibly a draft fragment. Unclear what tests this new density applies to, or how this test relates to the model released for community use. (update after later reading, now understand what this refers to, but still feel it was out of place and not well enough contextualized here)

Changed calibration to testing to make this clearer. The 1-layer approach is not included in the model released to the community.

Page 9 line 16 Soviet - capitalize.

Changed.

P 9 line 26 one OF the

Changed.

P9 – would be appropriate to acknowledge the lack of validation sites or validation data over Arctic sea ice, and uncertain accuracy of the products in that region.

We have added some comments regarding the uncertainty in precip earlier in the revised manuscript.

P11- Taking the median of the reanalysis products is an interesting idea if one has no idea which of the different products is best, but don't authors have better information about which is doing best from the comparison studies in literature?

Not really over Arctic sea ice. A new study of precipitation comparisons over the Arctic Ocean has recently been published (in early form) in the Journal of Climate but co-authors of this study, but the lack of validation data makes it challenging to recommend a particular product that one should focus on for such studies. This exercise was useful to guide us to exclude some reanalyses from our study - e.g. MERRA-2 - as stated in the MERRA data description.

P14 L10 – Initial conditions the Warren climatology is quite outdated. It is good you are trying to update them somehow. Is there evidence, e.g. from current autonomous ice mass balance buoys, that snow still regularly survives summer? Can you 'calibrate' this adjustment scheme based on those observations? Would a degree-day model be better than number of melting days? Also, what category is this snow placed in? Does it have a density reflective of melting snow (i.e. 400-500 kg/m3)?

Good question, and one that we can't answer with certainty about large-scale snow depth distributions in summer given the lack of observations. However, we can piece together information to hint at an answer. Based on these "pieces" (observations), some snow persists and the amount of snow that persists in summer has decreased relative to the Warren climatology.

In the IMB data, for example in 2012 and 2013, four IMB buoys show snow is present while six buoys show snow-free conditions in mid-to-late August. Based on survey line data from SHEBA and HOTRAX, snow can persist in summer where the largest drifts exist (next to ridges, etc.). However, when in the field, it's extremely difficult to tell the difference between a melting, slushy snow layer and a melting slushy ice surface (see the SHEBA field notes for an interesting reference for this – some scientists called this slush snow while others called it ice). What we can say with more confidence is that there is less snow in summer than there used to be based on the decreasing trend in surface albedo from AVHRR and melt pond information from MODIS data.

If the reviewer means a model based on the number of freezing-degree days, then our opinion is that the persistence of above-freezing days would have a larger effect on the removal of snow than the amount that's present in May.

We don't explicitly use a density in the initial conditions since we scale the climatological snow depth based on the number of consecutive above-freezing days.

P14 L22 – explain how this is 'linearly scaled' a bit better. Provide an equation. Is the fraction by which duration of melt is different from mean simply multiplied by snow depth? Does it mean that at 2x duration no snow is left and at 0x duration 2x snow is left?

We have provided more detail in the revised manuscript.

scaling factor = (mean duration for climatology - duration for individual summer)/mean duration for climatology

initial snow depth = august climatology*(1+scaling_factor)

We could add this to the manuscript if the reviewer thinks necessary.

*We only have ERA-Interim data for 1979-1991 rather than the full 1954-1991 climatological period to calculate the mean duration of above-freezing days in summer.

P14 L 28 – were necessary. . . Could this be because the model doesn't handle melt processes?

We don't believe so since deeper snow depths were needed to improve the comparison between the model and Soviet station observations (hence the initial conditions). Melt would reduce the snow depths so we don't think that is why.

P15Fig2 – These substantial August snow depths in 2012 and 2013 should be compared against available buoy data to determine if they are reasonable. The reviewer believes they are not and that this is ultimately a nonphysical tuning mechanism that helps account for lack of melt processes and poor representation of precipitation phase at this time of year in reanalyses.

The IMB data for 2012 and 2013 show four IMB buoys with snow present in summer (and six buoys with snow-free conditions). For the cases where snow is present, three of the four buoys show depths of ~10 cm and larger. The buoy IDs for persistent snow are: 2012C, 2012D, 2012G, and 2012I. Note, some of these buoys show data for summer 2013 and are not just limited to the year 2012, when they were deployed.

P 16 L 17 – this section is missing a clear statement of how accurate OIB data is expected to be.

The recent STOSIWIG study (Kwok et al., 2017) stated that snow depth can be estimated from a Snow Radar echogram with an uncertainty of 'several centimeters' although this depends strongly on the ice conditions, the particular Snow Radar system being used, and various other factors (e.g. geolocation errors associated with the plane pitch and roll. We have added a sentence regarding this to the discussion at the end of Section 3, where the OIB data are introduced.

P16 L 19-27 – pretty hand wavey – not rigorous.
Show plots of how snow evolves in model – what fraction of the snowpack is new snow layer vs time (it would have to be small to be realistic.)

We feel the plots already included in the manuscript provide the seasonal evolution of the budget terms highlighting the small contribution from the new snow layer.

P 17 Fig 3 – modeled data appears systemically low by about 5 cm depth. Snow density has essentially no relationship between modeled and observed. Individual year data – which is how this data would be used to derive altimetry based estimates of ice thickness – appear poor.

We didn't want to over fit the model to the data. Obviously it would be easy to calibrate this to increase snow depth in this comparison but we wanted to be flexible to the fact these depend strongly on the chosen forcings and the time period of analysis.

An r = 0.58 does not translate to no relationship. Clearly the snow depth shows a better relationship, but the density relationship isn't awful.

The validation part of the model is more in reference to the modern OIB data record, so we refer the reviewer to this.

P17 L 13 delete "extremely" – an error of ~5cm on a snowpack of ~20cm is still a 25% error.

We are referring to the seasonal cycle here and have attempted to make this clearer that we are referring to the seasonal cycle. We dropped extremely.

P18 L 1 – r of 0.74 would not generally be characterized as 'strong' P18 L2 - delete 'more' . . . its just moderate. Also, what is actually suggested here is that the model is good at predicting the MEAN – because you can tune your constants to make the mean look very nice, but not very good at capturing the interannual variability that is key to getting snow depth right for altimetry.

We are unsure how you have decided it's not good at getting the interannual variability? That's not really shown in this figure due to the lack of coincident data across years.

P18,L4-5 "In General, the moderate/high correlations. . . provide confidence. . ." This statement is hand waving and cheerleader- y without content. Delete this statement and replace it with a statement that articulates the degree of certainty with which output of the model should be treated. Suggest authors calculate the +-95% prediction interval for a modeled density or depth relative to this dataset. Suggest authors do this for individual locations/months on individual years, as well as mean.

At this stage we are just calibrating the model. The validation and expression of model errors comes in the OIB section.

P18, Figure 4 – This comparison is really just showing how well tuned the models are on average. Since the model is not presented as a mean climatology, but rather is presented as a deterministic snow product for specific locations on specific years, this comparison is inadequate.

Again, the OIB comparisons provide this later.

P19 L 3-6 Here authors make an odd argument. The model does not reproduce the climatology observations as well as a single mean over the entire timeframe. They argue this is OK because the model will handle interannual variability better because of its 'more advanced' density parameterization. The density parameterization is not particularly physically realistic, however, and fails to meaningfully capture interannual variability of the climatology density data (figure 3d). Reviewer therefore finds this statement lacking.

We state that the model is able to respond to expected interannual variability in the forcing, which is different to how the reviewer expresses this. Again we test this later with OIB.

We agree our approach is very simple, so have changed the more advanced line to read 'include a simple bulk density parameterization;

P19 l10 – not all marginal ice zones have low concentration – clarify that low concentration areas are where greater impact is expected

We have changed this to 'low ice concentration regimes'

P 19 L 19 – Reviewer disagrees that wind threshold velocity for blowing snow is un- constrained. Resources reviewed can establish that under all but extreme conditions (e.g. recent rain on snow) a threshold of 10 m/s is pretty high, maybe unreasonable. It would be better to range this within the values observed in the references cited earlier.

Changed this to poorly constrained. As you say this is on the high side, but we wanted o test the sensitivity so wanted to provide a big change, and doubling was consistent with the other changes. It does appear from this that increasing this value only slightly might lower the low bias in snow depths compared to the Soviet station data, but would also lower the density to create a low bias. This is obviously a tough balancing act and further calibration efforts will be explored in future.

P 20 L 15-18 – Speculative. Reviewer finds no reason to believe the median should be superior in regions of heavy snowfall. Defend or delete.

We have changed this to 'for simplicity'. As you say we provide no evidence that the results are better, but it is a useful synthesis forcing data to use - preventing the creation of figures across all the different forcing sets.

P21 L 5 – again this represents the mean over the decade being presented/compared. The model performance over this timeframe is highly tunable and not the performance metric of interest to an end user taking this data as an input to altimetry – that user would want to know the prediction interval for individual or moderate size groups of snow data points, and probably also whether there is any change in mean bias over time.

We have provided more information about the comparisons in terms of the root mean squared errors, which we feel are the best and most fair way of comparing these snow depth distributions.

P 22 Fig 6 – are standard deviations in depths this low comparable to any observations? If so they suggest a single climatology would be adequate for most end uses.

It's challenging to compare the SD from a pan-Arctic model with the SD from an in-situ dataset. We also dont have enough confidence in the snow density to make such a statement.

P22 L 15 plurals

Not sure what this refers to.

P23 L4 The fact that there is scatter among the reanalyses is not necessarily an argument for taking the median of them. Delete.

The point is that the snowfall forcings show strong differences so also having some kind of 'consensus' (e.g. median) snowfall dataset is useful. We have changed this line in the text.

P23 L19 providing significant VOLUME REDUCTIONS and sinks of snow (wind packing is not a sink)

Agreed, we have changed this to reductions of snow volume.

P24 L1 – Convergence really causes an increase in snow volume, not an increase in depth. The use of depth vs volume for a cell needs to be sorted out and treated consistently throughout this paper. Reviewer recalls a section way up at the top saying snow would be handled in volume throughout the paper, but has seen treatment vary.

We have made it clearer that we are tracking an effective snow depth (snow volume distributed across the entire grid-cell) and replaced volume with depth throughout the revised manuscript.

P25 Fig 9b – unclear what "Ocean" refers to. Snowfall directly into water? Caption needs to be more descriptive and the figure subcomponents should be linked back to which equation # they represent.

We have completed re-labeled this figure to improve clarity.

Fig 9f – this underscores the issue with treating convergence as a change in depth- in reality convergence/divergence of the ice at scale does not change depth in the sense that such would be used to interpret remote sensing. It changes snow volume, further, it appears the impact of convergence/divergence is noisy at best.

We treat this as a change in effective snow depth over the grid-cell. This is the most consistent way of treating this and avoids issues with the changing ice concentration.

Fig 9 I – density map warrants discussion. For example, density appears highest in central arctic (far from melt) This is likely untrue and an issue with not including melt/rain on snow processes. Very low density is indicated in marginal areas around N Greenland and in Baffin bay/Canadian islands. Can these be supported at all?

We agree that melt processes, especially in the marginal seas, should increase the density in these locations. This has been added as a future priority for the model.

Fig 9 h-k colors appear to fade toward lower value indication near land in general. Is this valid or a plotting artifact? Again volume and depth are used interchangeably and plotted on a depth scale. This needs to be resolved consistently.

Not sure I really see what the reviewer is referring to. Maybe some lower values in the older snow but likely due to wind packing as not seen in the new snow results..

Page 24 L 5-7– Authors state these measurements are explicitly for altimetry retrievals, so they must have characteristics useful for such, (more than matching seasonal evolution on aver- age) including: 1. Capture interannual variability 2. Capture spatial variability 3. No long term bias or trends in error (that could be mistaken as trends in ice thickness). If these cannot be shown, perhaps a discussion about whether this approach is viable is needed.

This all comes in the OIB validation section.

Page 26 Figure 10 – "As in figure 6" This is a nice addition to convey consistency, but pls also provide full description in caption, don't make reader hunt back several pages.

OK added.

Page 26 L15 Without melt processes included, what explains the loss in depth?

I believe in this case it is no significant accumulation and wind packing reducing the density. Can be shown in Figure S4.

Page 26 L 5 Why the depths are different during the later time period IN THIS MODEL is within the scope of demonstrating a model, even if understanding why they are changing in reality is not. Please answer: Are depths less due to less ice for snow to fall on? Or due to less precip? The reader must know if the model is representing the changing Arctic – since it is calibrated on old data. This cannot reasonably be scoped out of the study.

The snow depths are different between periods primarily due to the difference in the timing of sea ice freeze-up. Maximum snowfall rates occur in autumn (Warren et al., 1999; Webster et al., 2014; Lique et al., 2016; Boisvert et al., 2018), so if sea ice forms later, more snow falls into the open water relative to when sea ice formed earlier. This has been modeled in Blanchard et al. (2018) and Webster et al. (accepted), studies implementing a similar modeling approach to NESOSIM, and another study using a more sophisticated model (CCSM) (Hezel et al. 2012) as well as in ongoing work using CESM 1. This relationship has also been shown in Webster et al. (2014) based on ice mass balance buoy, *in situ*, airborne, and satellite data. The *in situ* and buoy data suggest no significant change in snowfall rates, while the combination of *in situ*, buoy, and airborne data show a decrease in snow depth that corresponds to later sea ice freeze-up (derived from passive microwave data).

We deliberately do not include a more in-depth discussion on this topic because 1) the manuscript is already at 13,000 words, and 2) we're preparing a manuscript on this topic and feel that including this discussion would detract from the purpose of the other manuscript. A comparison of the reanalysis precipitation between the 1980s and 2000s show little difference, which suggests that less precipitation (magnitude, phase) is not the primary driver of these differences. This is still open to debate, however. As the reanalysis description was very similar across the time periods we have now dropped this from the manuscript (following the recommendation of Reviewer 1). The 1980s budget figures have been moved to the SI too. Our aim is to focus on the model performance in the 2000s period, along with the OIB validation analysis. If the reviewer is interested in seeing more results on this topic, we would encourage him or her to contact Melinda Webster at melinda.a.webster@nasa.gov

P 27 Fig 11 – see comment on Figure 10

OK, changed.

P 29 Fig 12 – please clarify what positive and negative deviation mean (is the product higher or lower snow than the median product)?

Red (blue) colours indicate the individual reanalysis-forced simulations have higher (lower) snow depth.

P 30 line 3 – it is not clear that the median provides a result any more useful than the others. One should note which product compared best to coastal stations data and any other indications from literature which might be best.

The comparisons with coastal stations (e.g. Lindsay et al., 2014) are not wholly representative of precipitation biases over the Arctic Ocean. This is discussed more in Boisvert et al., (2018) and we reference this in the revised manuscript.

P 30 line 20 – this is not surprising and should be noted as such. Advection/convergence/divergence was much less important than snowfall in the plots above.

This is showing the spread based on the product spread, rather than its importance to the total mass balance. We agree this isn't surprising, but it is worth documenting and including.

P30 L 23-24 – here is where the idea of snow depth vs snow volume is really important. Dynamics are perhaps not important in depth over a 100km cell average, but they are important to the DEPTH on the actual subgrid ice, since divergence creates new ice with no snow, rather than rearranging all the snow into a gridcell average. The averaging over the 100km cell at each timestep may be particularly important in ice generating areas, where snow is continually averaged back into source regions, rather than being advected out entirely. Tracking ice classes within the cells, as is done in CICE may be critically important.

We agree this is worth thinking more about in future.

P 31 figure 13 – the drift scheme matters little over huge areas because convergence and divergence cancel. This plot is just not the right way to consider this, particularly in the context of use fore spatially distributed altimetry observations. Figure 14 suggests that the drift products don't differ that much between them in the central basin, but that having drift represented at all is very important, altering snowpack by O50% in large areas of the Arctic.

The point here is that if one cares about regional mean snow depths, the choice of product isn't that important really. Agree it matters more on the grid-scale hence the reason to show maps in the following figure.

P32 L6 – There are actually substantial biases in the peripheral seas – which may not be important overall, but cannot be ignored in the statement about biases.

Added ' and issues around the ice edge'

P33 – given the importance of concentration product, better understanding the role of changing concentration in the changing modeled snow depth above is important.

Agreed.

P34 – Tough to compare to observational data this noisy. Reviewer agrees they can be considered 'in agreement' within the bounds of the error of either. . . both of which are large. Are any of the OIB algorithms emerging as superior? Must all three be treated as equally likely?

Not sure noisy is the right phrase to use here, but there are clear differences in the products (an observational uncertainty). They seem to all have different pros and cons depending on the region/year/scale being analyzed. The STOSIWIG paper (Kwok et al., 2017) made no clear recommendations in this regard unfortunately.

P35 – comparison of 100km grid cells still includes substantial averaging, but already shows poor agreement. Agreement should be presented in terms of a 95 % prediction interval so user knows the capability of the method in useful terms – if the model says snow was xx OIB will say snow was xx +- yy 95% of the time.

Figure 14 shows the spread in snow depth across all the OIB campaigns, with the spread/uncertainty based on the product spread. If we wanted to add the individual product uncertainty we would have to guess at this as, they don't necessarily provide this. It is also thought to be highly variable and a function of the ice type and snow depth profiled. The spread is expected to be very large for a regional mean. The RMSE comparisons in Figure 15 etc. are a useful way of showing how the products and NESOSIM compare, but we acknowledge throughout the high uncertainty in both NESOSIM and OIB estimates making such comparisons challenging to interpret.

P36 – this discussion of the comparison of the scatter plots goes to great lengths to avoid describing the obvious. The model isn't very good at reproducing variability on OIB data, and if you believe OIB snow data is in any way representative of the variability in snow depth on ice, the modeled snow depth isn't very good at capturing spatial or interannual variability. The conclusion should then be that more sophisticated model representations are needed or that OIB data is trash. Since the model didn't agree with the Soviet drift station data scatter plot very well either, I don't think you can conclude that the model is adequate but OIB is trash.

We are confused by this statement. We very clearly present the correlations and rmse values for the model and OIB comparisons, then also show the mean interannual variability comparisons. The reader can clearly make their own opinion regarding the comparison but I don't think that would be that either the model or OIB are trash. The correlations are moderate/strong in general, there are no obvious skews/biases and the rmse is ~10 cm, with some values lower than this. Both the model and OIB have uncertainties associated so saying stronger statements than we have should be done with caution. Hopefully our inclusion of the mW99 comparisons help put thse in context.

P37 – "in general, however, the moderate to strong correlations. . . gives us confidence" Reviewer cussed in exasperation when reading this. This is a science paper not an opinion piece. These are not moderate to strong correlations! They clearly show NE- SOSIM cannot capture the variability observed well. Get this subjective language out of the paper and replace it with quantifications of how well the model does at both rep- resenting means (where performance is good because of tuning) and variability (where the model is not working so well). Talk about whether the model is good enough to be used in altimetry honestly and present some paths forward to getting there if it isn't.

See discussion above. We did indeed include an interpretation of the comparison metrics we presented in this study based on the chosen metrics, which is pretty standard practice. The choice of moderate/good/strong was based on standard definitions for the interpretation of correlation coefficient values.

P 37 L 19 – data is yet to be released in parenthesis? Thin its out now. . .

We changed this to 'the data was not available for this study but was made available during the review phase of this paper'

P 37 L 20 – There must be some field data available that you could at least spot check it against!

Unsure what the benefit of a spot-check would be in this instance as it wouldn't be a particularly robust comparison.

P 37, L 26 delete very strong, delete good

We would like to keep these in. The agreement with the mean seasonal cycle was very good (very high correlation).

P 37 L 28 contributing to the MODELED seasonal evolution in snow depth

Added.

P38 L5 uncapitalize New, consider replacing with 'more recent.'

New Arctic was cited earlier so we wish to keep this statement.

P38 L7 There is no evidence presented that this median product is better, and good reason to believe it just averages in erroneous values and non-physically jumps be- tween atmospheric states toward limited representation of extreme events. Defend the use on scientific merit or consider deleting the median product.

We dropped the last part of this sentence.

P38 L10 use consistent language. . . it is 2nd order on mean, but first order in some regions.

Added 'in our regional mean analysis'

P 38 L 14 "moderate/strong correlations" This statement is flatly unsupported by the results shown, and authors 'confidence' in line 16 is unfounded. The product does not represent the OIB data well in terms of the intended use – in retrieving thickness from freeboard.

We are discussing correlation coefficients above 0.5, with some above 0.6/0.7. Moderate/strong are the appropriate way of describing these values based on the statistical guidelines we referred to.

Please provide a variable list

Thank you. We have added a variable list to the paper.

The NASA Eulerian Snow on Sea Ice Model (NESOSIM) v1.0: Initial model development and analysis

Alek A. Petty[1,2], Melinda Webster[1], Linette Boisvert[1,2], Thorsten Markus[1]

[1] Cryospheric Sciences Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA.
[2] Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD, USA.

Correspondence to: Alek A. Petty (alek.a.petty@nasa.gov)

Abstract. The NASA Eulerian Snow On Sea Ice Model (NESOSIM) is a new open source model that currently produces daily estimates of the depth and density of snow on sea ice across the Arctic Ocean through the accumulation season~~polar oceans~~. NESOSIM has been developed in a three-dimensional Eulerian framework and includes two (vertical) snow layers and several simple parameterizations (accumulation, wind packing, advection/divergence, blowing snow lost to leads) to represent ~~the~~ key sources and sinks of snow on sea ice. The model is forced with daily inputs of snowfall and near-surface winds (from reanalyses), sea ice concentration (from satellite passive microwave data) and sea ice drift (from satellite feature tracking), during the accumulation season (August through April). In this study, we present the NESOSIM formulation, ~~initial~~ calibration efforts, sensitivity studies and validation efforts across an Arctic Ocean domain (100 km horizontal resolution). The simulated snow depth and density are calibrated with in-situ data collected on drifting ice stations during the 1980s. NESOSIM ~~demonstrates~~ shows ~~very~~ strong agreement with the in-situ seasonal cycles of snow depth and density, and shows good (moderate) agreement with the regional snow depth (density) distributions. NESOSIM is run for a contemporary period (2000 to 2015), with t~~T~~he results ~~exhibit~~ showing strong sensitivity to the reanalysis-derived snowfall forcing data, with the MERRA/JRA-55 (ASR) derived snow depths generally higher (lower) than ERA-Interim. We also ~~derive~~ generate and force NESOSIM with a ~~new~~ consensus 'median' daily snowfall dataset from these ~~three~~ reanalyses~~reanalysis datasets to improve reliability in our input snowfall data~~. ~~NESOSIM is run for a contemporary period (2000 to 2015) and~~ The results are compared~~compared~~ against snow depth estimates derived from NASA's Operation IceBridge (OIB) S~~s~~now R~~r~~adar data from 2009-2015, showing moderate/strong ~~agreement~~correlations ~~, especially in the 2012-2015 comparisons~~and root mean squared errors of ~10 cm depending on the OIB snow depth product analyzed, ~~.~~similar to the comparisons between OIB snow depths and the commonly used modified Warren snow depth climatology. Potential improvements to this initial NESOSIM formulation are discussed in the hopes of improving the accuracy and reliability of these simulated snow depths and densities.

1 Introduction

Snow on sea ice is a crucial component of the polar climate system. Its low thermal conductivity modulates sea ice growth through the cold winter months (e.g. Maykut and Untersteiner, 1971, Sturm et al., 2002), while its high surface albedo limits solar radiation absorption and thus inhibits sea ice melt in spring and summer (e.g. Warren, 1982; Grenfell and Perovich, 1984; Perovich, 2002). Conversely, freshwater production from snow melt facilitates melt pond formation in spring/summer which lowers the surface albedo and promotes sea ice melt (Eicken et al., 2002; 2004). The accumulation of snow on sea ice also modulates the freshwater flux into the ocean, a key component of the freshwater budget of the Arctic (e.g. Serreze et al., 2006).

Estimates of snow depth on sea ice are also a required input for deriving sea ice thickness from satellite altimetry, e.g. from ESA's CryoSat-2 (e.g. Laxon et al., 2013) and NASA's upcoming ICESat-2 mission (Markus et al., 2017). The altimetry technique involves measurements of sea ice freeboard, the extension of sea ice above a local sea level, and estimates of snow depth to derive sea ice thickness, with snow depth being one of the primary sources of uncertainty for both laser and radar altimetry (e.g. Giles et al., 2007). Poor knowledge of snow density provides a further source of uncertainty through its influence on the ice freeboard and radar penetration into the snow pack (e.g. Giles et al., 2007, Kern et al., 2015).

Unfortunately, direct observations of snow depth and density across the polar oceans are ~~lacking~~very limited, due to difficulties in remotely sensing this relatively thin (O(*10* cm)) and heterogeneous medium, and logistical challenges associated with in-situ data collection. Passive microwave data have been used to estimate snow depth over first-year ice on a basin-scale across both poles (e.g., Markus and Cavalieri 1998, Comiso et al., 2003, Maass et al., 2015), although these data are arguably more relevant for the first-year dominated Antarctic sea ice pack and tend to underestimate snow depth in deformed sea ice regimes (e.g. Worby et al., 2008; Brucker and Markus, 2013). Combinations of satellite and/or airborne sensors with variable snow penetration depths are also being explored as a means of producing basin-scale snow depth estimates (e.g. Armitage and Ridout, 2015, Guerreiro et al., 2016; Kwok and Markus, 2017), although this approach is still in its infancy and has limited temporal coverage. NASA's Operation IceBridge has provided airborne measurements of snow depth on sea ice since ~~its launch in~~ 2009 (Kurtz et al., 2013). However, the Arctic snow depth data collected are primarily limited to the western Arctic sea ice cover in spring (the spring 2017 campaign also included a flight over the eastern Arctic Ocean), while the Southern Ocean data have only been briefly explored to-date (e.g. Kwok and Maksym, 2014). For the Arctic, a climatology of snow depth produced from Soviet drifting station data collected prior to

1991 (Warren et al., 1999) is still commonly used as a basin-scale snow depth product. ~~As such, the sea ice community often relies on simple models of snow depth forced by reanalyses (primarily snowfall data) (e.g., Maksym and Markus 2008; Kwok and Cunningham, 2008; Blanchard-Wigglesworth et al., 2018) or, for the Arctic, a climatology of snow depth produced from Soviet drifting station data collected prior to 1991 (Warren et al., 1999).~~ The Soviet drifting station data also provide the only observationally-based basin-scale assessment of snow density currently available. This snow climatology is ~~also~~ expected to be outdated due to the rapid changes experienced in the Arctic climate system over the last few decades (Webster et al., 2014), although recent efforts have been made to modify this climatology based on ice type (halving the climatology over first year ice, e.g. Laxon et al., 2013, Kwok and Cunningham, 2015).~~.~~

Due to these observational limitations, the sea ice community often utilize simple models of snow depth forced by reanalyses (primarily snowfall data) (e.g., Maksym and Markus 2008; Kwok and Cunningham, 2008; Blanchard-Wrigglesworth et al., 2018). More sophisticated snow on sea ice models are available, such as SnowModel, a terrestrial snow model recently adapted for sea ice environments (Liston et al., 2018), as well as the prognostic snow layer included in sea ice climate model components, such as CICE (Hunke & Lipscomb, 2010) and the Louvain-la-Neuve Sea Ice Model (LIM) which have recently undergone various improvements to their snow physics (Holland et al., 2011; Lecomte et al., 2015).

In this study we present a new model to derive snow depth (and density) across the Arctic Ocean. Our aim is to develop a model of physical and computational simplicity to allow for a detailed assessment of the sensitivity of snow depths to the various input forcing data needed to produce seasonal, basin-scale, snow depths. The spread in reanalysis-derived snowfall estimates over the Arctic Ocean is high (Boisvert et al., 2018), while the importance and uncertainty of other forcing data (e.g. ice concentration and drift) are still largely unknown. We also wanted a model that could be forced with observed ice concentration and drift to help accurately constrain the seasonal sea ice cycle - a challenge for the more sophisticated sea ice models described above.

~~In this study we present a new model to derive snow depth (and density) across the polar oceans. Our objective~~ Our overall goal is that NESOSIM can be used to ~~is to~~ produce reliable basin-scale daily snow depth and density estimates needed for satellite altimetry calculations of sea ice thickness for both historical analyses and near real-time operations across the polar oceans. We thus expect the model to increase in complexity with future model developments, e.g. new parameterizations or improvements to existing parameterizations as needed. A secondary utility of the model will be the production of daily/monthly/seasonal snow depths from reanalysis data that can help guide climate modelling research efforts addressing the representation and importance of snow on sea ice in the global climate system.

In the following sections, we present and describe the model configuration/physics, the sensitivity of the model to the input forcing data (e.g. reanalyses snowfall, satellite-derived ice drifts), and ~~initial~~ model calibration/validation efforts. We focus this initial study solely on the Arctic, however our plan is for the model to be applied and tested in a Southern Ocean framework in the near future. We conclude by looking ahead to potential improvements in the model physics and planned future activities related to our efforts to improve our understanding of snow on sea ice.

2 Model description

The NASA Eulerian Snow On Sea Ice Model (NESOSIM) is a three-dimensional, two-layer (vertical), Eulerian snow budget model developed with the primary aim of producing daily estimates of snow depth and density across the polar oceans. NESOSIM includes several parameterizations that represent key mechanisms of snow variability through the snow accumulation/growth season, and two snow layers to broadly represent the evolution of both old/compacted snow and new/fresh snow. The model schematic is shown in Figure 1. ~~Our aim was to produce a model of physical and computational simplicity to allow for a detailed assessment of the sensitivity of the modeled snow depths to the various input data used. We expect the model to increase in complexity with future model developments, e.g. new parameterizations or improvements to existing parameterizations as needed. We decided on a Eulerian snow budget approach (as opposed to a Lagrangian approach, e.g. Kwok and Cunningham, 2008) for a number of reasons: (i) it provides a framework flexible to the presence (or lack of) ice drift data, increasing the utility of the model in regions/time periods where ice drift data might be lacking, (ii) it provides a simple assessment of the spatial significance of the parameterized budget terms included in the model, including ice dynamics, and (iii) the parameterizations developed in this framework can be easily transferred to other Eulerian sea ice models (e.g. the Los Alamos sea ice model CICE) included in General Circulation Models (GCMs). The following subsections detail the model setup and various parameterizations currently included in NESOSIM.~~

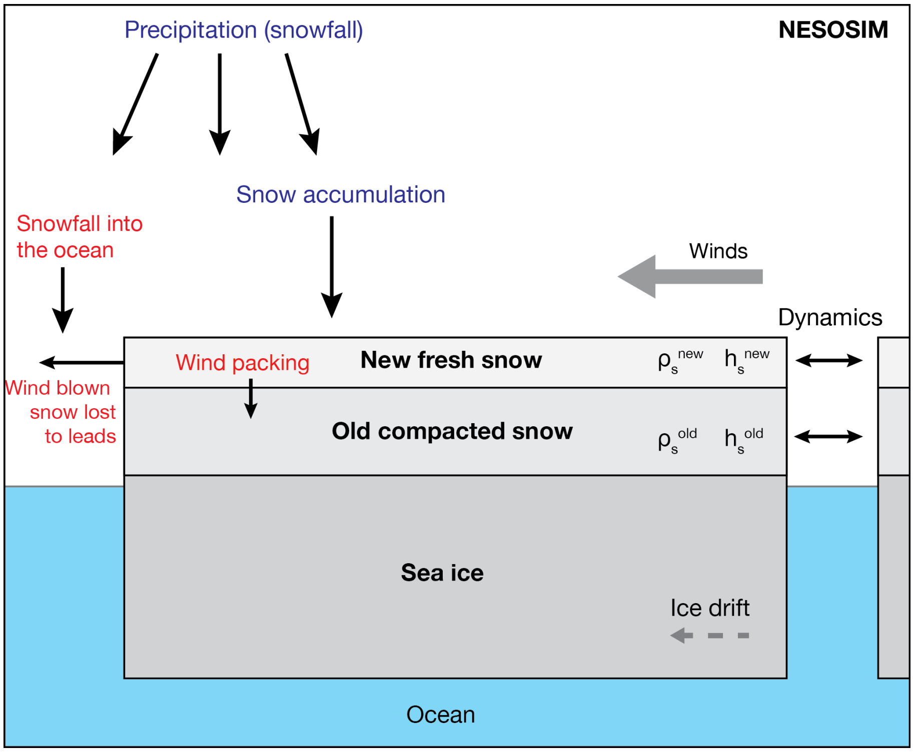


Figure 1: Schematic of the NASA Eulerian Snow On Sea Ice Model (NESOSIM) presented in this study. The red (blue) text indicates processes that result in a loss (gain) of snow. 'Dynamics' indicates the combination of ice/snow advection and convergence/divergence which can cause either loss or gain of snow.

We decided on a Eulerian snow budget approach (as opposed to a Lagrangian approach, e.g. Kwok and Cunningham, 2008) for a number of reasons: (i) it provides a framework flexible to the availability (or lack of) ice drift data, increasing the utility of the model in regions/time periods where ice drift data might be lacking, (ii) it provides a simple assessment of the spatial significance of the parameterized budget terms included in the model, including ice dynamics, and (iii) the parameterizations developed in this framework can be easily transferred to other Eulerian sea ice models (e.g. the sea ice climate model component CICE) included in General Circulation Models (GCMs). The following subsections detail the model setup and various parameterizations currently included in NESOSIM.

2.1 Model configuration

NESOSIM includes two vertical layers on an x/y horizontal grid, with each horizontal grid-cell and snow layer featuring a prognostic snow ~~volume~~ depth and fixed snow density. This two-layer approach was taken to represent the strong differences in properties between dense snow, associated with wind slab, and fresh ~~, cold~~

snow from recent snowfall, while keeping the model computationally efficient and the model physics easily trackable. As stated previously, our plan is that NESOSIM will be used for studying snow on sea ice across the Arctic and Southern Oceans, however, for this initial analysis we run the model on a 100 km x 100 km polar stereographic grid covering the Arctic Ocean and peripheral seas (~~shown~~ model domain shown later in Figure 5). This grid resolution was chosen due to considerations of computational efficiency and the horizontal resolutions of the various input data.

The model is forced with daily data of snowfall and near-surface winds from reanalysis data, satellite passive microwave ice concentration, and satellite-derived ice drifts~~The model is forced with daily data of snowfall, near-surface winds, ice concentration, and ice drift~~. ~~We discuss the forcing datasets used in this study in Section 3. Note that in the model we track the evolution of snow volume (the volume of snow per unit grid cell, in units of meters) for simplicity, instead of snow depth. An effective snow depth within a given grid cell is then produced by dividing the snow volume by the grid-cell ice concentration. The model run is initiated each year from the end of summer (default of August 15th, rationale discussed in Section 2.5) to the middle of spring (May 1st), to avoid the complexity of snow melt processes expected outside of these dates. We hope to extend the model runs into the melt season in future model development efforts. The default model configuration (forcings/parameter settings) are given in Table 1.~~

Model variable	
Snow accumulation (m), h_s^{acc}	
Snowfall into the ocean (m), S_f^{oce}	
Snow dynamics (m), h_s^{dyn}	
Snow divergence (m), h_s^{div}	
Snow advection (m), h_s^{adv}	
Wind packing (m), h_s^{wp}	
Blowing snow lost to leads (m), h_s^{bs}	
Effective snow depth (m), h_s	
Physical snow depth (m), h_s/A	
Bulk snow density (kg m^{-3}), ρ_s^b	
Model parameter	**Default setting**
New snow density (kg m^{-3}), ρ_s^n	200
Old snow density (kg m^{-3}), ρ_s^o	350

Wind packing coefficient (s^{-1}), α	5.8 x 10^{-7}~~0.05~~
Blowing snow coefficient (m^{-1}), β	2.9 x 10^{-7}~~0.0025~~
Wind ~~packing~~ action threshold (m s^{-1}),	5
Forcing data	
Snowfall (kg m^{-2}), S_f	ERA-I/MEDIAN-SF (as specified)
Near-surface winds (m s^{-1}), U	ERA-I
Sea ice concentration, A	Bootstrap
Sea ice motion (m s^{-1}), u_i~~drift~~	NSIDCv3 (Polar Pathfinder)

Table 1: Default model forcings and parameter settings used by NESOSIM.

At each daily time step, an effective snow depth within each grid cell is produced from our various snow budget terms (described in the following subsections) using a forward Euler method as

Eq. (1):

$$h_s(t+1,0,x,y) = h_s(t,0,x,y) + \Delta h_s^{acc}(t,x,y) + \Delta h_s^{dyn}(t,0,x,y) + \Delta h_s^{wp}(t,0,x,y) + \Delta h_s^{bs}(t,x,y)$$

and

$$h_s(t+1,1,x,y) = h_s(t,1,x,y) + \Delta h_s^{dyn}(t,1,x,y) + \Delta h_s^{wp}(t,1,x,y).$$

where t denotes the daily time index, the second index indicates the relevant snow layer (0 = new, 1 = old), and x and y are the horizontal grid indices. ~~Our modelled snow~~NESOSIM ~~is partitioned into~~ uses two vertical ~~density~~ layers: a “new” layer, $\rho_s{}^n$, which represents recent snowfall, and an “old” layer, $\rho_s{}^o$, which represents snow that has undergone wind compaction and snow grain metamorphism (Colbeck, 1982; Sturm and Massom, 2017). These two fixed snow densities are justified in more detail in the following subsection. We track the evolution of an effective snow depth within each grid-cell (the volume of snow per unit grid cell area) for simplicity. The actual snow depth over the ice fraction is calculated by dividing the effective grid-cell snow depth by the grid-cell ice concentration.

We also calculate a bulk snow density, which is the weighted average density across the two snow layers, as

Eq. (2):

$$\rho_s^{b}(t,x,y) = ((h_s(t,0,x,y)\rho_s^{n} + h_s(t,1,x,y)\rho_s^{o})/(h_s(t,0,x,y) + h_s(t,1,x,y))).$$

Note that the bulk snow density is masked if the respective ice concentration in the given grid-cell is less than 15%, or the effective snow depth is less than 2 cm. While the model tracks the snow budget terms for all grid cell ice concentrations, only grid-cells with an ice concentration above 15% are shown in the analysis, to prevent spurious interpretations in regions of near open water conditions.

Each annual model run is initialized in the middle of summer (default of August 15th, rationale discussed in Section 2.5) and run until the following spring (May 1st). This early summer start time was chosen to include the significant snowfall expected across the Central Arctic through August (Radionov et al., 1997; Warren et al., 1999; Boisvert et al., 2018) while also avoiding the periods of significant snow melt in late spring/early-mid summer. We acknowledge that this end of August time period still likely includes surface melt events that are not captured/included in this model, but are hoped to be addressed in future model developments. We also apply a variable initial snow depth (at t = 0) across our model domain, as discussed in Section 3.4. New ice that subsequently forms in a given grid-cell is assumed to be snow free.

2.2 Snow accumulation

To accumulate snow ~~in our model~~on a given grid-cell, the snowfall water equivalent from our reanalysis ~~data~~ field is converted to snow ~~volume~~ depth using a representative snow density. Snow pit and density data from the Surface Heat Budget of the Arctic Ocean (SHEBA) experiment and the Soviet drifting ice station data helped guide the parameterization of our seasonal snow density evolution. ~~Our modelled snow is partitioned into two density layers: a "new" layer, which represents recent snowfall, and an "old" layer, which represents snow that has undergone wind compaction and snow grain metamorphism (Colbeck, 1982; Sturm and Massom, 2017).~~ Initially, snow accumulates into the new/fresh snow layer within a given grid cell as

Eq. (3~~1~~):

$$\Delta h_s^{acc}(t,x,y) = S_f(t,x,y)\, A(t,x,y)/\rho_s^{n}\,,$$

where S_f (in units of kg m^{-2}) is the gridded daily snowfall across the model domain~~, ρ_s^{n} is the density of the new snow layer~~ and~~,~~ A is the gridded daily ice concentration~~, t is the daily time index, and x and y are the horizontal grid indices~~. ~~The density of the new snow layer is fixed at ρ_s^{n} = 200 kg m^{-3}. This value implicitly represents a~~

~~combination of cold, dry snowfall (~150 kg m^{-3}) and wet snowfall (~230 kg m^{-3}) based on direct observations over Arctic sea ice (Radionov et al., 1997; Sturm et al., 2002).~~ The density of the new snow layer is fixed at $\rho_s{}^n$ = 200 kg m^{-3}. This value implicitly represents a combination of cold, dry snowfall (~150 kg m^{-3}) and wet snowfall (~230 kg m^{-3}) based on direct observations over Arctic sea ice (Radionov et al., 1997; Sturm et al., 2002).

Snow can be transferred from the new snow layer to the old snow layer depending on the strength of the near-surface wind forcing. The old snow layer is an implicit combination of two layers that, on average, comprise the majority of the snowpack bulk ~~density~~mass: wind slab and depth hoar (Sturm et al., 2002; Sturm, 2009). The density of wind slab ranges between ~300 kg m^{-3} and ~41~~0~~0 kg m^{-3} on average (Colbeck, 1982; Radionov et al., 1997; Warren et al., 1999; Sturm et al., 2002), while depth hoar has an average density of ~150 - 250 kg m^{-3} (Colbeck, 1982; Sturm et al., 2002). Based on SHEBA data, ~~both layers contribute roughly equally to the bulk thickness of~~ the Arctic snow cover consists of slightly more wind slab than depth hoar, comprising ~80% of it collectively (Sturm et al., 2002). For this reason, we use ~~the~~ a weighted average of the higher-end values of wind slab and depth hoar as the density value for the old snow layer, $\rho_s{}^o$. However, we note that the density and ratio of wind slab and depth hoar layers depends on several factors including the atmospheric conditions during precipitation events, sea ice surface roughness, snow depth, and the internal snowpack temperature gradient (Sturm et al., 2002). We did experiment with alternative snow densities (e.g. the wider spread of 150 and 400 kg m^{-3}) but found this provided worse correspondence with the seasonal snow density evolution compiled from in-situ Soviet station data (introduced in Section 3.4). ~~Our simple parameterization scheme is thus expected to be generally representative of basin-wide conditions, but will contribute to uncertainty in our modeled snow depths.~~

When wind speeds are greater than 5 m s^{-1}, the change in snow depth from wind packing between the two snow layers respectively is given as:

Eq. (2):

$$\Delta\, h_s{}^{wp}(t,0,x,y) = -\alpha\, T_d\, h_s(t,0,x,y) \text{ for } U(t,x,y) > \omega$$

$$\Delta\, h_s{}^{wp}(t,1,x,y) = (\rho_s{}^n/\rho_s{}^o)\alpha T_d h_s \cancel{h_s}(t,0,x,y) \text{ for } U(t,x,y) > \omega$$

where $U\cancel{(x,y)}$ is the 10 m wind speed, ω is a wind ~~speed~~ action threshold for wind packing to occur (default of 5 m/s), α is a wind packing coefficient which determines the fraction of the new snow layer that is transferred into the old snow layer (default value of 5.8 x $10^{-7\cancel{8}}$ s^{-1}) and~~,~~ T_d is the number of seconds in our~~a~~ daily time step~~y~~ (= 86400 s)~~, and $\rho_s{}^{0}$ is the density of the old snow layer~~. The second grid index in Eq. 2 (values of 0 and 1) represents the vertical snow layers. The wind action threshold of 5 m s^{-1} was determined based on observational and modeling studies of blowing snow in the terrestrial Arctic and sea ice environments (Pomeroy et al., 1997; Radionov et al., 1997; Sturm and Stuefer, 2013), while the wind packing coefficient is a free/unconstrained parameter in the model.~~.~~

2.3 Ice/snow dynamics

Snow within a given grid cell can also evolve due to ice ~~drift~~motion. Here we adapt the ice concentration budget approach used in e.g. Holland and Kimura (2016) (and more recently in Petty et al., 2018) to snow ~~volume~~ depth as

Eq. (3):

$$\Delta\, h_s{}^{dyn}(t,x,y) = -\nabla.\,(h_s(t,x,y)\cancel{.}\,u_i(t,x,y)),$$

where u_i is the daily gridded ice ~~drift~~motion. As in the ice concentration budget studies discussed above, we can expand this into a divergence/convergence term, and an advection term, as

Eq. (4):

$$\Delta\, h_s{}^{div}(t,x,y) = -h_s(t,x,y)\,.\,\nabla(u_i(t,x,y)) \text{ and}$$

$$\Delta\, h_s{}^{adv}(t,x,y) = -\nabla(h_s(t,x,y))\,.\,u_i(t,x,y)\,,$$

where $\Delta\, h_s{}^{div}$ is the change in effective snow ~~volume~~ depth from divergence/convergence, i.e. changes due to spatial gradients in ice ~~drift~~motion, and $\Delta\, h_s{}^{adv}$ is the change in snow depth ~~volume~~ from advection, i.e. changes due to spatial gradients in snow depth ~~volume~~ (assuming constant drift). Note that this parameter is applied to both 'old' and 'new' snow layers concurrently.

2.4 Blowing snow lost to leads

Snow within a grid cell can also be lost to leads/open water in the ice pack due to the impact of wind forcing, i.e. blowing snow lost to leads. This parameter is expected to be most significant in regions where high lead fractions, wind speeds and snowfall (e.g. the marginal ice zone in the North Atlantic sector of the Arctic) are expected to result in significant wind blown snow lost to leads/open water (e.g. Leonard and Maksym, 2011). Note that we only apply this wind loss term to the new snow layer as we assume the 'old' wind packed snow layer is immune to the impact of wind forcing (e.g. Petrich et al., 2012; Trujillo et al., 2016). The blowing snow to leads term is calculated as

Eq. (5):

$$\Delta\, {h_s}^{bs}(t,x,y) = -\beta\, T_d U(t,x,y)\, h_s(t,0,x,y)(1 - A(t,x,y)) \text{ for } U(t,x,y) > \omega_s \cancel{,}$$

where β is a blowing snow coefficient (default value of 2~~5~~.9~~8~~ x $10^{-7\cancel{8}}$ m^{-1} ~~s^{-1}~~). This is a free/unconstrained parameter in the model, with its default value chosen through our model calibration efforts.

We also keep track of snow that enters the ocean through snowfall into the open water fraction and blowing snow lost to leads, a quantity of relevance to those interested in the freshwater budgets of the polar oceans. This is given as

Eq. (6):

$${S_f}^{oce}(t,x,y) = S_f(t,x,y)\,(1 - A(t,x,y))/{\rho_s}^n \;- \cancel{+}\, \Delta\, {h_s}^{bs}(t,x,y)\,.$$

~~2.5 Model evolution~~

~~At each daily time step, the snow volume and density within each grid cell is updated based on the budget terms described above using a forward Euler method as~~

~~Eq. (7):~~

$$\cancel{h_s(t+1,0,x,y) = h_s(t,0,x,y) + \Delta\, {h_s}^{acc}(t,x,y) + \Delta\, {h_s}^{dyn}(t,0,x,y) + \Delta\, {h_s}^{wp}(0,x,y) + \Delta\, {h_s}^{bs}(x,y)}$$

~~and~~

~~$h_s(t+1,1,x,y) = h_s(t,1,x,y) + \Delta h_s^{dyn}(t,1,x,y) + \Delta h_s^{wp}(1,x,y)$.~~

~~Note that we also calculate a bulk snow density, which is the weighted average density across the two snow layers, as~~

~~Eq. (8):~~

~~$\rho_s^{b}(t,x,y) = ((h_s(t,0,x,y)\rho_s^{n} + h_s(t,1,x,y)\rho_s^{o})/(h_s(t,0,x,y) + h_s(t,1,x,y))$.~~

~~As discussed earlier, each model run is initialized at the end of summer (default of August 15th) and run until the following spring (May 1st). This early summer start time was chosen to include the significant snowfall expected across the Central Arctic through August (Radionov et al., 1997; Warren et al., 1999; Boisvert et al., submitted). We acknowledge that this end of August time period also likely includes surface melt events that are not captured/included in this model but are hoped to be addressed in future model developments. We also apply a variable initial snow volume (at $t = 0$) across our model domain, as discussed in the following section. The model is easily adaptable, such that it can be reinitialized from a later date and run forward in time, making it suitable for near real-time operations, an expected use of this model in the near future.~~

For ~~initial~~ model ~~calibration~~ testing we also ran NESOSIM with different combinations of the model parameterizations discussed above. When we turn off the wind-packing parameterization, snow remains fixed in the 'new' snow layer, despite the strength of the wind forcing, so the model effectively becomes a 1-layer model. To account for the low bias in snow density expected by constraining the snow density to the density of fresh/new snow, we forced this snow layer with the daily climatological snow density based on Warren et al., (1999), which we refer to as ρ-W99.

3 Model forcing and calibration/validation data

In the following subsections we describe the forcing data and calibration/validation data used in this study, including atmospheric forcing data (snowfall and winds), satellite-derived ice motion~~drifts~~, satellite-derived ice concentration, S~~s~~oviet drifting station snow depths/densities (for model calibration) and Operation IceBridge snow depths (for model validation).

3.1 Atmospheric forcing

We use snowfall data provided by the European Center for Medium Range Weather Forecast (ECMWF) ERA-Interim (ERA-I) reanalysis. ERA-I is a global reanalysis that utilizes a 4D variational data assimilation scheme (Dee et al., 2011). We use the 12-hourly ERA-I snowfall data from August 15th 1980 to May 1st 1991 and August 15th 2000 to May 1st 2015. We use the 0.75° x 0.75 ° horizontal resolution data, which are summed to produce daily snowfall estimates across the Arctic. ERA-I snowfall data have been used in previous studies exploring snow accumulation over Arctic sea ice (e.g. Kwok and Cunningham, 2008; Blanchard-Wrigglesworth et al., 2018)~~Blanchard-Wigglesworth et al., 2018)~~, while comparisons of reanalysis-derived precipitation data with coastal weather stations suggests ERA-I is one of the better products available for Arctic studies (Serreze and Hurst, 2000; Lindsay et al., 2014). A more detailed comparison of snowfall/precipitation estimates over the Arctic Ocean has recently been carried out alongside this study (Boisvert et al., 2018)~~(Boisvert et al., submitted)~~, which we expect to build on in the future.

Reanalysis	**Producer**	**Resolution***	**Coverage**
ERA-Interim	European Centre for Medium-Range Weather Forecasts (ECMWF)	0.75° x 0.75°	1979 - present (NRT, few months data latency)
ASRv1	Various contributors, see Bromwich et al., (2016)	30 km x 30 km	2000 - 2012
JRA-55	Japanese Meteorological Agency (JMA)	0.56° x 0.56°	1958 - present (NRT, few months data latency)
MERRA	NASA's Global Modeling and Assimilation Office (GMAO)	0.5° x 0.66°	1979 - June 2016

Table 2: Summary of the four different reanalysis datasets used in this study (data availability often subject to change/updates, information given at the date of submission). NRT: Near real-time. *different resolutions available in some cases.

We explore the sensitivity of our results to the input snowfall data by forcing the model with snowfall estimates provided by three additional reanalysis-derived snowfall products. Unfortunately, not all reanalyses provide direct estimates of snowfall (and rainfall), and instead provide just total precipitation, e.g. the data from the widely used National Centers for Environmental Prediction (NCEP)-National Center for Atmospheric Research (NCAR) Reanalysis 1 and 2, so we focus our analysis on three other commonly used reanalyses that provide direct estimates of snowfall: the Japanese Meteorological Agency 55-year Japanese reanalysis (JRA-55); NASA's Modern-Era Retrospective Analysis for Research and Application (MERRA); and the Arctic System Reanalysis, version 1 (ASRv1), as described below and summarized in Table 2.

JRA-55: The Japanese Meteorological Agency (JRA) 55-year Japanese reanalysis (JRA-55) is a global atmospheric reanalysis that utilizes a 4-D variational assimilation system covering the period 1958 to present (Kobayashi et al, 2015). JRA-55 was developed as an improvement to their previous 25-year reanalysis (JRA-25), which we do not include in this study. We use the daily JRA-55 snowfall data from August 15th 1980 to May 1st 1991 and August 15th 2000 to May 1st 2015. The data were obtained from the National Center for Atmospheric Research's Research Data Archive at a horizontal resolution of 0.56° x 0.56° (~ 60 km), downscaled from the original 1.25 ° x 1.25 ° Gaussian grid. The data are being produced on a near real-time basis (2-6 month data latency).

*MERRA***:** NASA's Modern-Era Retrospective Analysis for Research and Application (MERRA) is a global reanalysis that utilizes a 3D variational data assimilation scheme within the Goddard Earth Observing System Data Assimilation System (GEOS-5) (Rienecker et al, 2011). We use the daily MERRA snowfall data from August 15th 1980 to May 1st 1991 and August 15th 2000 to May 1st 2015. The data are provided at a horizontal resolution of 0.5° (latitude) by 0.66 ° (longitude). Note that an updated version of MERRA (MERRA-2) is also available, but is known to have a high precipitation bias compared to the other reanalyses (Boisvert et al., 2018) ~~(Boisvert et al., submitted)~~ so we exclude this from our study.

ASRv1: The Arctic System Reanalysis, version 1 (ASRv1) is a regional reanalysis based on the Weather Research and Forecasting model (Polar WRF) that utilizes a 3D variational data assimilation scheme and is adapted for the polar regions (Hines and Bromwich, 2008). The ASRv1 data are only available from 2000 to 2012, so we use the daily snowfall data from August 15th 2000 to May 1st 2012, which is provided at a horizontal resolution of 30 km x 30 km.

Considering the expected importance and uncertainty of the reanalysis-derived snowfall for deriving snow depth, we also produce a synthesized snowfall dataset by taking the median snowfall across the gridded snowfall products, for each daily grid-cell (data referred to as MEDIAN-SF). We use the gridded ERA-I, JRA-55 and MERRA snowfall data, as these products all cover the longer-term (1980-2015) time period.

NESOSIM also requires daily estimates of near-surface winds to drive the wind packing and wind loss terms, which we take from the ERA-I reanalysis for all reanalysis model runs. Jakobsen et al., (2012, Figure 2) show that ERA-I winds had the lowest biases of several reanalysis-derived near-surface wind estimates compared to TARA drifting station data. We compute the magnitude of the winds from the six-hourly *u/v* vectors before averaging to produce a daily (gridded) wind magnitude dataset.

We linearly interpolate all the daily snowfall (and ERA-I wind magnitude) estimates onto our 100 km x 100 km polar stereographic model domain. Gridding scripts written in Python are included in the GitHub code repository.

3.2 Satellite derived ice ~~drift~~ motion data

We primarily make use of the daily Polar Pathfinder ice ~~drift~~ motion data, version 3 (Tschudi et al., 2016) made available through the National Snow and Ice Data Center (the product is referred to herein as NSIDCv3). A daily ice ~~drift~~ motion vector is calculated using a cross-correlation technique applied to sequential daily satellite images acquired by passive microwave satellite sensors (i.e. a one day lag in parcel tracking) which are blended via optimal interpolation with estimates from the International Arctic Buoy Programme (IABP) and wind data from the National Centers for Environmental Prediction (NCEP)-National Center for Atmospheric Research (NCAR) Reanalysis. The data are available daily from October 1978 through February 2017 (at the time of writing) at a horizontal resolution of 25 km x 25 km. In this study we use the daily data from August 15th 1980 to May 1st 1991 and August 15th 2000 to May 1st 2015. We grid the daily ice ~~drift~~ motion data onto our 100 km model domain (using linear interpolation) and smooth the data using a simple Gaussian filter (as in Holland and Kimura, 2016 and Petty et al., 2018).

Recent studies have explored the uncertainty in satellite-derived ice ~~drift~~ motion data (Sumata et al., 2014) and errors introduced by the NSIDC interpolation methodology (Szanyi et al, 2016). We thus also explore the sensitivity of the model results to the input ice ~~drift~~ motion data by forcing the model with ice ~~drift~~ motion estimates provided by three additional satellite-derived ice ~~drift~~ motion products, as described below and summarized in Table 3.

Product	Resolution	Daily lag	Data source	Coverage	Availability

NSIDCv3	25 km	1 day	AVHRR, SMMR, SSM/I , AMSR-E,IAPBs, NCEP-R1	October 1978 - Feb 2017	Public
OSI-SAF	62.5 km	2 days	ASCAT*	October 2010 - present	Public/NRT
KIMURA	60 km	1 day	AMSR-E, AMSR-2	Jan 2003 - Sep 2011 / July 2012 - Dec 2016	On request
CERSAT	62.5 km	3 days	ASCAT*	January 2007 - present?	Public/NRT

Table 3: Summary of the different ice ~~drift~~ motion datasets used in this study based on information obtained at the time of submission. *These agencies produce drift datasets using different individual/combinations of satellite sensors not utilized in this study. NRT: Near real-time.

OSISAF: The European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) produce a number of low-resolution sea ice ~~drift~~ motion products from satellite passive microwave sensors and scatterometry (Lavergne, 2010). Here we use the merged ice ~~drift~~ motion product, which increases coverage and reliability over their single sensor drift products (Lavergne, 2010). The merged drift product uses a 2 day lag in ice parcel tracking and a Continuous Maximum Cross Correlation (CMMC) method to optimize the drift product, and is available daily (October through April) since 2010 at a horizontal resolution of 62.5 km x 62.5 km.

CERSAT: The Centre ERS d'Archivage et de Traitement (CERSAT), part of the Institut Français de Recherché pour l'Exploitation de la Mer (IFREMER) produce a number of ice ~~drift~~ motion datasets by merging various combinations of satellite passive microwave and scatterometry data (Girard-Ardhiun and Ezraty 2012). Here we use data produced from the merging of Advanced Scatterometer (ASCAT) and the Special Sensor Microwave Imager (SSMI) data, which are available daily (September to May) since 2007 at a horizontal resolution of 62.5 km x 62.5 km. Note that CERSAT provide data using both a 3 and 6 day lag in the tracking of ice displacement, but we use the 3 day lag data as this is closest to the 1 day lag used by the NSIDCv3 product.

KIMURA: The KIMURA drift data are produced using brightness temperatures obtained by the Advanced Microwave Scanning Radiometer for EOS (AMSR-E) from January 2003 to September 2011 and the Advanced Microwave Scanning Radiometer 2 (AMSR-2) from July 2012 to December 2016 using a cross-correlation approach (see Kimura et al., 2013 for more details). Wintertime (November-December, January-March) ice ~~drifts~~ motion vectors are derived using the 36-GHz channel, while the summertime drifts used in this study (August-October, April) are derived using the 18-GHz channel, to maximize the reliability and coverage of the data. The data are provided at a 60 km x 60 km horizontal resolution.

We use data from these three additional products from August 15th 2010, 2012, 2013 and 2014 to May 1st of the subsequent years, a period of coincident data coverage across the four drift products (including NSIDCv3). We linearly interpolate all the daily drift datasets onto the 100 km x 100 km polar stereographic model domain used in this study. As highlighted above, not all the products produce drift estimates in August, or even September, so for those products we assume no ice ~~drifts~~ motion through this period. To investigate the importance of ice ~~drift~~motion, we also run the model assuming no ice ~~drift~~ motion for the entire model simulation (NODRIFT), as discussed in more detail later.

3.3 Sea ice concentration

We use the daily Bootstrap sea ice concentration (SIC) data, version 3 (Comiso, 2000 updated 2017), which are produced from passive microwave brightness temperature estimates and made available through the NSIDC. We choose to primarily use the Bootstrap over, for example NASA Team data (Cavalieri et al., 1996 updated 2017), another commonly used SIC dataset, as Bootstrap SIC data are less sensitive to surface melt, producing higher concentrations in general. We use the NASA Team data in a sensitivity study to explore the sensitivity of the model to this choice of sea ice concentration data. Due to differences in satellite orbit and sensor characteristics, the SIC data feature a time-varying pole hole depending on the passive microwave sensor used. As we require consistent SIC data across the pole hole, we follow the approach of Petty et al., (2018) and apply a mean SIC calculated in a 0.5^{o} halo around the variable pole hole to all grid cells within the pole hole. The data are provided at a 25 km x 25 km resolution polar stereographic grid from 1978 through 2016, and we use the daily data from August 15^{th} 1980 to May 1^{st} 1991 and August 15^{th} 2000 to May 1^{st} 2015. We linearly interpolate the daily SIC data onto our 100 km x 100 km model domain. Note that a gap in the passive microwave record exists from December 3rd 1987 to January 13th 1988, so we do not run the model through the 1987-1988 winter period.

3.4 Soviet station data and initial conditions

We use *in-situ* snow data collected on the former Soviet Union's drifting ice stations for initial model calibration and to help guide our choice of initial conditions (Radionov et al., 1997; Warren et al., 1999; Fetterer and Radionov, 2000). The drifting ice stations were in operation in 1937 and 1954-1991, although in this study we use the field observations collected from 1980-1991 due to the temporal overlap with the model forcing data. During the drifting ice stations, snow depth data were collected every 10 days in 10 m intervals along a 500 m or 1000 m survey line. Snow density measurements were made every ~100 m along the same survey lines, and atmospheric conditions were recorded at near-daily frequencies. Despite their limited spatial coverage, these data provide the most complete record of snow and atmospheric conditions to date over the Arctic sea ice pack.

Initial conditions: We initialize the model on August 15th of each year with a snow depth representing the fraction of snow assumed to have survived the summer melt season and/or accumulated during summer. The August snow depth climatology compiled by Warren et al., (1999, referred to herein as W99) from the Soviet station data suggests significant amounts of snow (up to 10 cm) are present in late summer, especially over the Central Arctic sea ice north of Greenland (Radionov et al., 1997). This inclusion of an initial snow depth was also guided by our preliminary model calibration studies that showed that including these initial conditions provided a better match with the seasonal snow depths observations (calibrations presented later). To produce initial mid-August snow depths, we use a near-surface air temperature-based scaling of the August W99 snow depth climatology to account for changes in the duration of the summer melt season (e.g. Markus et al., 2009). Briefly, we calculate the annual number of days with continuous, above-freezing, air temperatures (taken from the ERA-I reanalysis), which we refer to here as the summer melt duration. To create an initial (August) snow depth estimate for a given year, we linearly scale the W99 August snow depth climatology based on the summer melt duration of the chosen year and the climatological summer melt duration given in Radionov et al., (1997). If the melt duration is longer than the climatological mean in a specific region, the scaled August climatology reflects a reduction in snow depth in August due to the longer melt season. The snow depth is then distributed evenly over the 'old' and 'new' snow layers based on the climatological observations that some snow persists through summer (Radionov et al., 1997), and the occurrence of summer snowfall events (Radionov et al., 1997; Perovich et al., 2017). While admittedly this is a crude approach for parameterizing an initial snow depth, our sensitivity studies demonstrated that initial conditions were necessary to improve the comparison with the drifting station observations (as presented and discussed in the following section), and indicate that late summer snowfall events might play a significant role in establishing the snow cover on Arctic sea ice prior to the fall/winter season (Warren et al., 1999). The August W99 snow depth climatology and temperature scaled initial snow depth estimates (for 2012 and 2013) are shown in Figure 2.

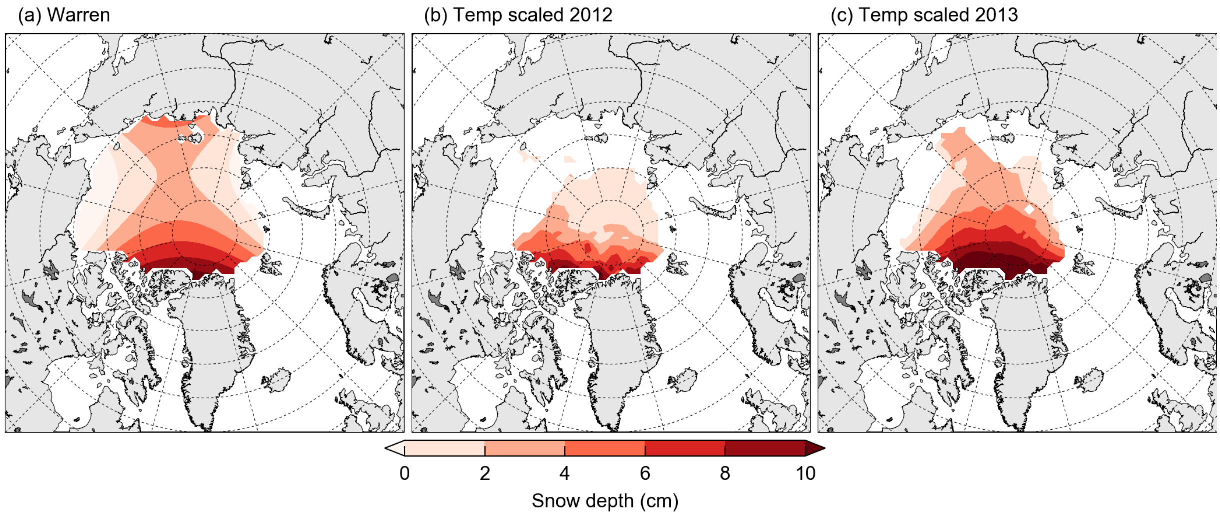


Figure 2: (left) Warren climatology of August snow depth, (middle and left) the initial conditions used in this study (broadly representing the snow ~~volume~~ depth as of August 15th) for 2012 and 2013 respectively, calculated using near-surface air temperature scaling.

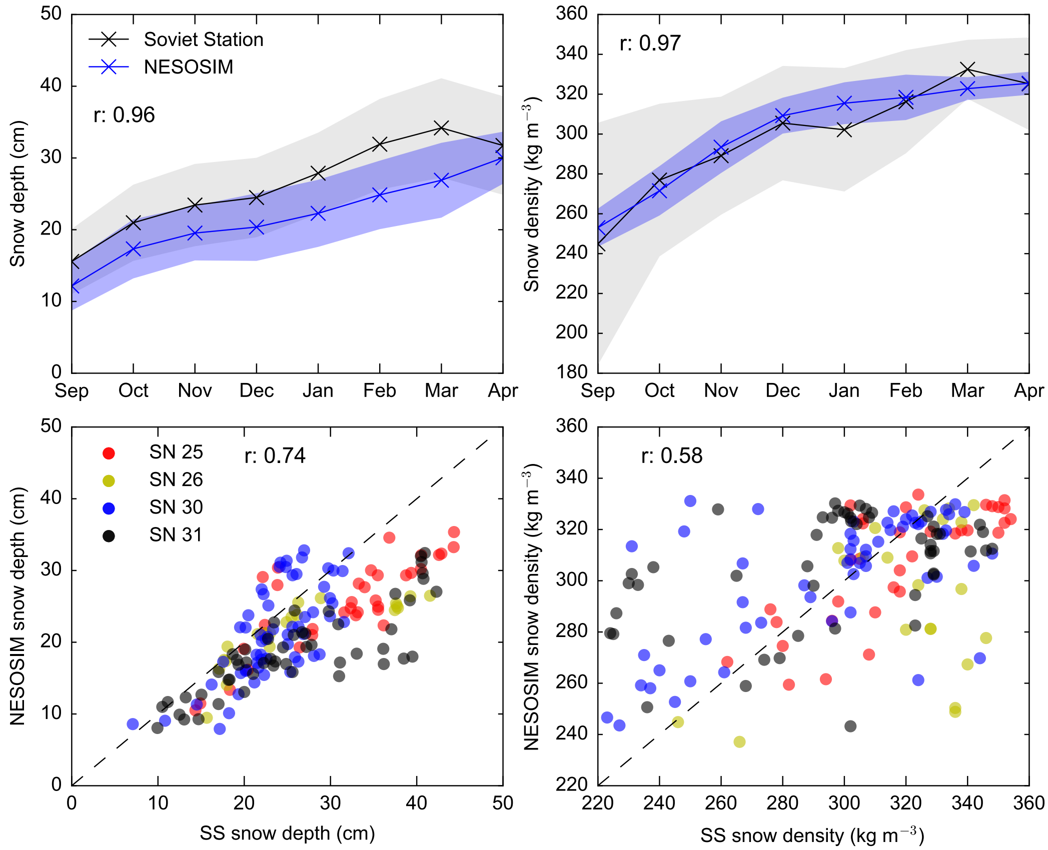
Model calibration: For our model calibration we use the raw snow depth and density data from the Soviet drifting stations 25, 26, 30 and 31. The data represent the average of a given survey line. The majority of survey lines remained constant each time they were sampled, so the dataset is a near-continuous time-series with a 10-day temporal resolution. Most survey lines were 1000 m in length, although in the earlier part of the historical record (e.g., before the 1980s), some ice stations had survey lines that were 500 m in length. Maps of the drifting stations are given in the supplementary information (Figure S3). Briefly, Station 25 drifted from the Central Arctic to the East Siberian Sea providing data from autumn 1981 to spring 1984, Station 26 drifted around the north of the East Siberian Sea providing data from autumn 1983 to spring 1984, Station 30 drifted around the north of the East Siberian Sea providing data from autumn 1988 to winter 1991, and Station 30 drifted around the Beaufort Sea providing data from winter 1989 to winter 1991. We use a simple nearest neighbor algorithm to match the data to the nearest model grid-cell for the relevant day the drifting station data were collected.

3.5 NASA's Operation IceBridge data

We compare our NESOSIM snow depth estimates with spring snow depths collected by NASA's Operation IceBridge (OIB) airborne mission. NASA's OIB mission began collecting airborne observations of the polar regions in 2009, bridging the gap between NASA's Ice, Cloud, and land Elevation Satellite (ICESat) mission which retired in 2009, and the future ICESat-2 mission scheduled for launch in the summer of 2018 (Markus et al., 2017). The OIB aircraft carry a suite of instruments designed to measure both land and sea ice, including their overlying snow cover. Here we primarily make use of snow depth estimates derived from the ultra-wideband Snow Radar (Panzer et al., 2013), which are available at a 40 m along-track resolution. These snow depths are thought to carry an uncertainty of several centimeters, although this depends strongly on the ice/snow conditions, the particular Snow Radar system being used, and various other factors, e.g. geolocation errors associated with the plane pitch and roll (e.g. Kurtz et al., 2013; Kwok et al., 2017 and references therein). Various algorithms have been developed to produce snow depth estimates from the OIB Snow Radar data (Kwok et al., 2017), with the products showing broad agreement in the regional snow depth distributions, but significant intraregional and interannual differences, due primarily to changes in the radar configuration and algorithm tuning. To account for these differences we use the snow depth data from the (i) Snow Radar Layer Detection (SRLD) (Koenig et al., 2016), (ii) NASA Goddard Space Flight Center (GSFC) (Kurtz et al., 2013) and (iii) Jet Propulsion Laboratory (JPL) (Kwok and Maksym, 2014; Kwok et al., 2017) snow depth products, that have produced, and made available, snow depth estimates at a 40 m along-track resolution from 2009 to 2015. We bin the 40 m OIB snow depth data onto our 100 km model grid and keep only the grid cells that included a significant quantity (> 1000 points) of the raw snow depth data. The OIB data are provided for various days through spring of the relevant campaign (data from mid-March to early-May, depending on the campaign year), so we grid the OIB data daily, and compare this with coincident (daily) NESOSIM snow depth estimates. The OIB data are collected mainly over the western Arctic sea ice, limiting our validation efforts to this region of the Arctic.

4 Model calibration and analysis

We carried out ~~initial~~ model calibration over the period Aug 15th 1980 - May 1st 1991 due to the coincident Soviet station data available during this period. As noted previously, this excludes the 1987-1988 winter season due to the lack of complete sea ice concentration data available during this period. As stated earlier, our initial calibration efforts involved manually tuning NESOSIM to improve the general fit with the mean seasonal snow depth/density cycles shown in the Soviet station data. Specifically, we included the temperature-scaled initial August snow depths and tuned both the wind packing coefficient, α (Eq. 5), and blowing snow coefficient, β (Eq. 6). We decided against a more optimized calibration effort due to limitations in the calibration data, i.e. its sparse availability in space/time and differences in spatial scales. We instead used the Soviet ~~drifting~~ station data to guide our model choices to achieve a more realistic seasonal cycle in snow depth and density. We also decided against specific model configuration parameter tuning due to these limitations in the calibration data, however this should be considered when analyzing the model performance, especially with regard to our validation efforts (i.e. more sophisticated and/or configuration specific tuning could improve the comparisons shown).


Soviet Station
NESOSIM
r: 0.96
Snow depth (cm)
r: 0.97
Snow density (kg m^{-3})
Sep
Oct
Nov
Dec
Jan
Feb
Mar
Apr
SN 25
SN 26
SN 30
SN 31
r: 0.74
NESOSIM snow depth (cm)
SS snow depth (cm)
r: 0.58
NESOSIM snow density (kg m^{-3})
SS snow density (kg m^{-3})

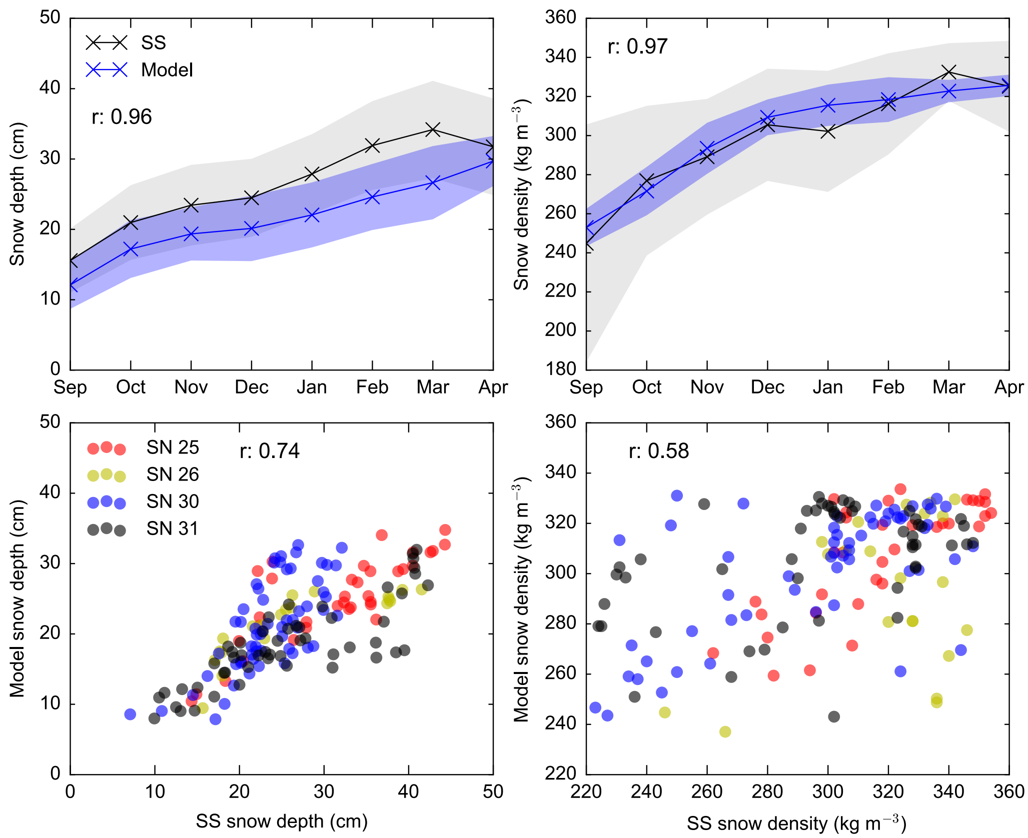


Figure ~~3:~~3: Comparison of NESOSIM snow depth (left) and snow density (right) data with drifting Soviet station data collected between 1981 and 1991. The top panels show the mean seasonal evolution of the snow depth and density for the model (blue) and Soviet station data (black), with the data binned into the different months the data were collected. The shaded area represents one standard deviation from the annual monthly mean. The bottom panels show scatter plots of all points for which there were temporal crossovers. The r-values indicate the correlation coefficient, while the colors indicate the different stations that collected the data. The NESOSIM data are from the default/ERA-I model configuration.

In Figure 3 we show comparisons of our NESOSIM results using the default model configuration (summarized in Table 1) and ERA-I snowfall forcing with the drifting station snow depth and density data. Figure 3 shows both the mean seasonal cycle based on all drifting station data points and coincident model grid cell values over this time period binned monthly, and the correlations of snow depth and snow density for all coincident data (described in Section 3.4). Our calibrated NESOSIM results ~~capture extremely well~~agree well with the mean ~~seasonal drifting station~~ seasonal cycle in snow depth (r = 0.96 with a low bias of ~3 to 7 cm) and snow density (r = 0.97, no significant seasonal bias) in the drifting station data. The large spread in the in-situ snow density in September-October is due to the survival of snow through the summer melt season (high density) and recent autumn snowfall (low density). The correlations between the raw drifting station data and NESOSIM snow depths are lower, but still strong (r = 0.74), while the snow density correlation strength is ~~more~~ moderate (r = 0.58), suggesting the model may be better capturing regional variability in snow depth over snow density. It should be noted, however, that snow density is highly variable in space and subject to large measurement uncertainties when collected in situ (Sturm, 2009). In general, the moderate/high correlations and seasonal comparisons provide confidence in the utility of NESOSIM for estimating snow depths across the Arctic.

In Figure 4, we highlight the sensitivity of NESOSIM to the chosen model configuration/sophistication, broadly representing the heuristic model tuning that was undertaken. First we tested the results of NESOSIM with different combinations of the various model parameterizations included. Note that as discussed at the end of Section 2, when we turn off the wind-packing parameterization the model essentially becomes a one layer model so we use a fixed Warren et al., (1999) seasonal snow density climatology (constant density value across the Arctic). As this is based on the same drifting station data we compare our results to, it is perhaps unsurprising that this configuration provides a better match with the seasonal drifting station snow depth cycle, including deeper snow depths (and reduced low snow depth bias) from November to April. We chose to develop NESOSIM to allow for the production of snow depths that agree well with the old drifting station snow climatology, but able to also respond to the expected interannual variability and trends in Arctic climate over recent decades, hence the decision to develop and include a simple bulk density parameterization.

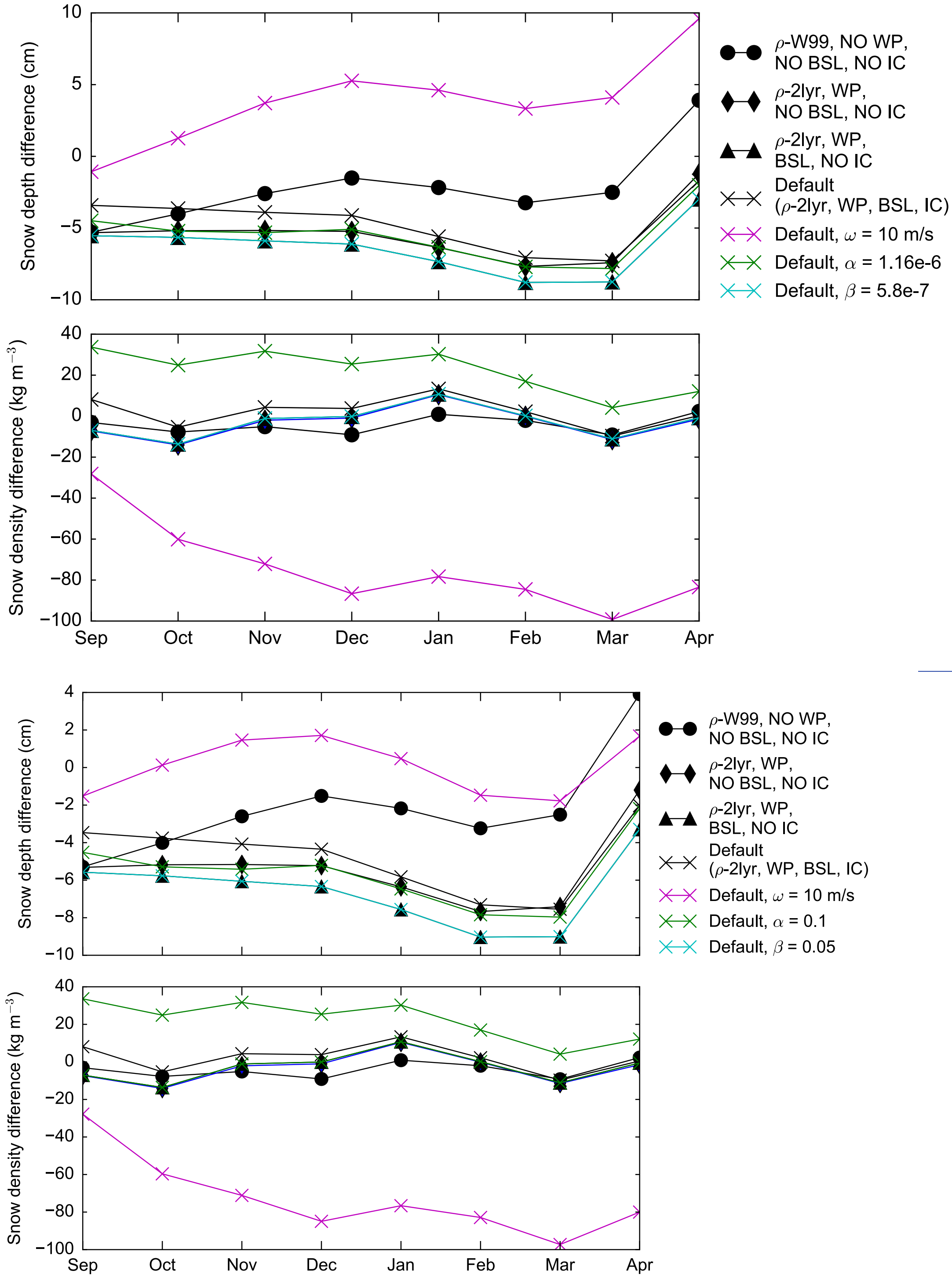


Figure 4: Differences between the mean (1980-1991) seasonal cycles in the drifting station data against various configurations of NESOSIM. The different symbols represent different levels of model sophistication, ρ-W99:

climatological Warren snow density, ρ-2lyr: default prognostic two-layer snow density, WP: wind packing parameterization, BSL: blowing snow loss parameterization, IC: initial conditions. NO indicates that the parameterization/initial conditions have been turned off. The different colours then represent a doubling of individual model parameters, with all other settings fixed to the default settings (see Table 1). The black crosses/line represents the default/ERA-I results (as shown in Figure 3). ~~UPDATE ALPHA AND BETA VALUES~~

~~In Figure 4, we highlight the sensitivity of NESOSIM to the chosen model configuration/sophistication, broadly representing the heuristic model tuning that was undertaken. First we tested the results of NESOSIM with different combinations of the various model parameterizations included. Note that as discussed at the end of Section 2, when we turn off the wind-packing parameterization the model essentially becomes a one layer model so we use a fixed Warren et al., (1999) seasonal snow density climatology (constant density value across the Arctic). As this is based on the same drifting station data we compare our results to, it is perhaps unsurprising that this configuration provides a better match with the seasonal drifting station snow depth cycle, including deeper snow depths (and reduced low snow depth bias) from November to April. We chose to develop NESOSIM to allow for the production of snow depths that agree well with the old drifting station snow climatology, but able to also respond to the expected interannual variability and trends in Arctic climate over recent decades, hence the decision to develop and include more advanced density parameterizations.~~

Including the blowing snow loss parameterization resulted in slightly lower snow depths (~2 cm), but no significant change in snow density. This parameterization can impact the bulk density implicitly by reducing the amount of fresh snow contributing to the total snow depth/density. As the drifting station data are collected primarily within the Central Arctic where ice concentrations are near to 100%, it was expected that including blowing snow loss would not result in significant differences, as this parameterization is expected to provide more of an impact in ~~the marginal ice zone~~lower ice concentration regimes, where unfortunately in-situ snow depth data are lacking. Including the initial snow depths resulted in a small increase in snow depth and density, especially earlier in the seasonal cycle, as expected, reducing the low bias compared to the drifting station data. The seasonal correlations were similarly high across these model configurations, highlighting the primary role of the model configuration choices in determining the general bias of the seasonal snow depth/density cycle.

As a simple demonstration of the sensitivity of the model to the ~~relatively un~~poorly constrained/unconstrained model parameters introduced in NESOSIM (the wind packing threshold, ω, the wind packing coefficient, α, the blowing snow loss coefficient, β), Figure 4 also shows results from NESOSIM with these three model parameters individually doubled (based on the default/ERA-I configuration). Doubling the wind packing threshold, ω, (from

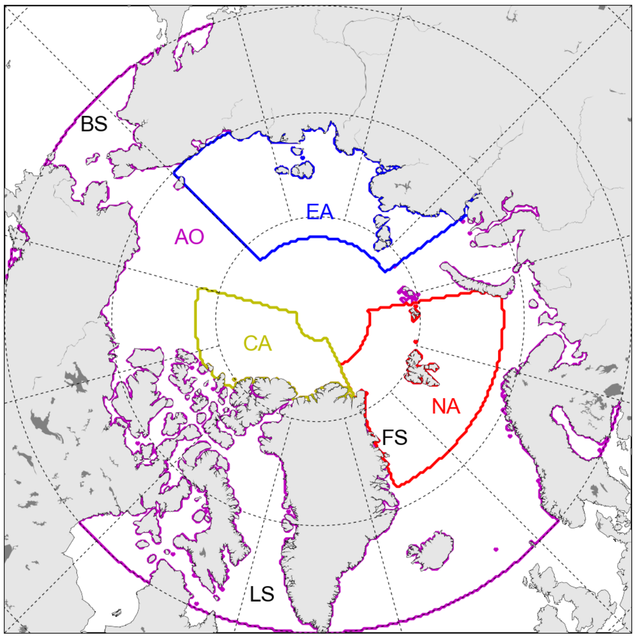
5 to 10 m/s) has a large impact on both the snow depth and density. By essentially reducing the likelihood for wind packing to occur, the snow accumulates and remains in the fresher 'new' snow layer for longer, significantly reducing the bulk snow density and increasing the seasonal snow depths. While this does produce snow depths that appear to agree ~~well~~ better with the drifting station data, the low bias in the seasonal snow density suggest this is unphysical. Doubling the wind packing coefficient, α, (from 5.8e-7 to 1.16e-6) ~~(from 0.05 to 0.1)~~ has broadly the opposite effect, as expected, reducing the snow depths by increasing the transfer of snow from the fresher 'new' snow layer to the denser 'old' snow layer. Doubling the blowing snow loss coefficient, β, (from 2.9e-7 to 5.8e-7~~0.025 to 0.05~~) has a negligible impact, again likely due to the location of the in-situ data away from the lower concentration ice regimes where this process is more significant.

As stated earlier, the differences in spatial scales and data coverage (time and space) make interpreting these comparisons/calibrations challenging. Specific model configurations may be required based on user demands, and our expectations is for these calibrations to evolve as new calibration data are made available and physical parameterizations introduced/updated. Note that we also compared the simulations of NESOSIM forced by the MERRA and JRA-55 snowfall data (Figures S2 and S3 provided in the Supplementary Information). In general the seasonal correlations with the drifting station data were similar to the ERA-I results, but the correlations of the raw data were slightly lower for JRA-55 (r = 0.69 for snow depth and r = 0.58 for snow density) and significantly lower for MERRA (r = 0.44 for snow depth and r = 0.57 for snow density). As discussed in Section 3, it is likely that specific model configuration tuning could improve these comparisons and the later validation efforts, but we decided against a more optimized calibration approach due to the limitations in the Soviet station data.

As discussed in Section 3.1, we also produced a synthesis snowfall dataset (MEDIAN-SF) using the median snowfall across the gridded ERA-I, JRA-55 and MERRA datasets. The MEDIAN-SF forced results are similar to the ERA-I results (Figures S4), in general, and show correlations similar to ERA-I and JRA-55 (r = 0.68 for snow depth and r = 0.58 for snow density). The MEDIAN-SF seasonal snow depths have a reduced low bias compared to the ERA-I results, although this difference is small. For the rest of this analysis we choose to mainly focus on the MEDIAN-SF forced results using the default configuration (Table 1) ~~as we expect these results to be less prone to errors in the individual reanalyses and more reliable in regions/periods of challenging (e.g. heavy) snowfall~~for simplicity. We provide a further assessment of the impact of the snowfall data in the following regional analysis and when we analyze the regional distributions across the more recent (2000-2015) time period.

~~4.1 Regional analysis~~

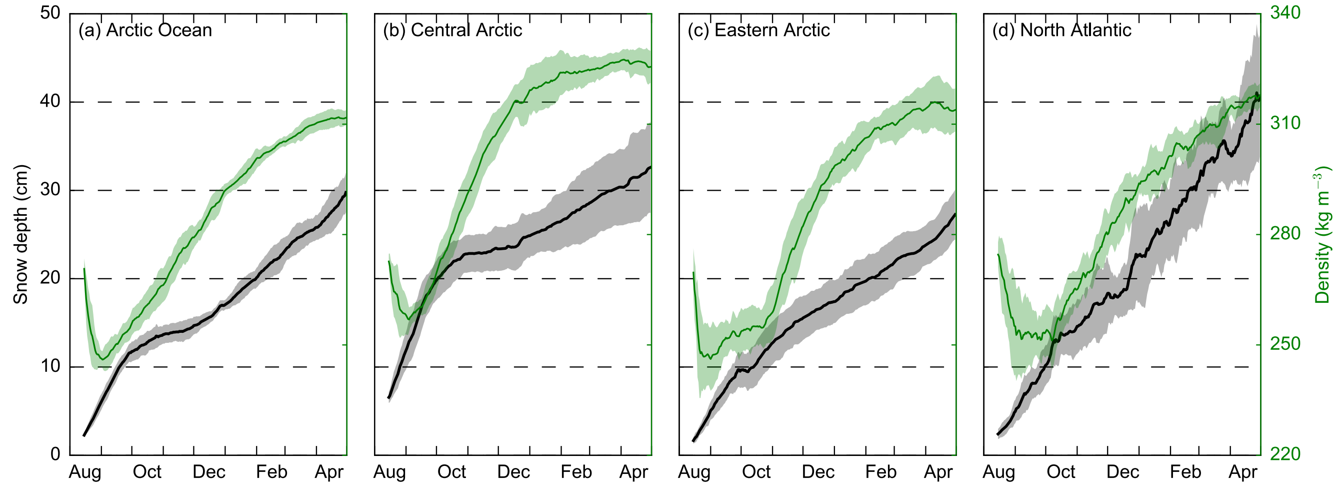
~~Here we provide a more detailed assessment of the regional NESOSIM results during this early Soviet station period (1980-1991). We focus our analysis on the Arctic Ocean (AO, everything north of 60 °N) and three specific regions that were chosen to represent different components of the Arctic sea ice/climate system: (i) the Central Arctic (CA, captures the thicker/multi-year ice over north of Greenland), (ii) the Eastern Arctic (EA, the increasingly first-year ice dominated sea ice regime), (iii) North Atlantic (NA, a region influenced by the transpolar ice drift and the North Atlantic storm track), as shown in Figure 5.~~



~~**Figure 5**: Map of the Arctic model domain and regions used in this study: AO: Arctic Ocean, CA: Central Arctic, EA: Eastern Arctic, NA: North Atlantic. BS: Bering Sea, LS: Labrador Sea are peripheral seas discussed in the manuscript.~~

~~Figure 6 shows the seasonal snow depth and density evolution across our four study regions for the 1980-1991 time period, using the default/MEDIAN-SF configuration (Table 1). The AO and CA region especially show strong initial increases in snow depth through fall (August to October) with the snow depth increasing at a slower rate from November to May, which is in good agreement with the W99 climatology. The EA and NA regions show a more uniform increase in snow depth from August to April. The NA region shows more daily snow depth variability, which was expected due to the strong ice drifts and the location of the NA storm track where passing~~

~~cyclones can deposit large quantities of snow in a short period of time. By May 1st the mean snow depths (and interannual variability, calculated as one standard deviation of the annual values) are given as: 29.8 +/- 2.2 cm (AO), 32.6 +/- 5.2 cm (CA), 27.3 +/- 2.8 cm (EA), 40.7 +/- 6.9 cm (NA).~~

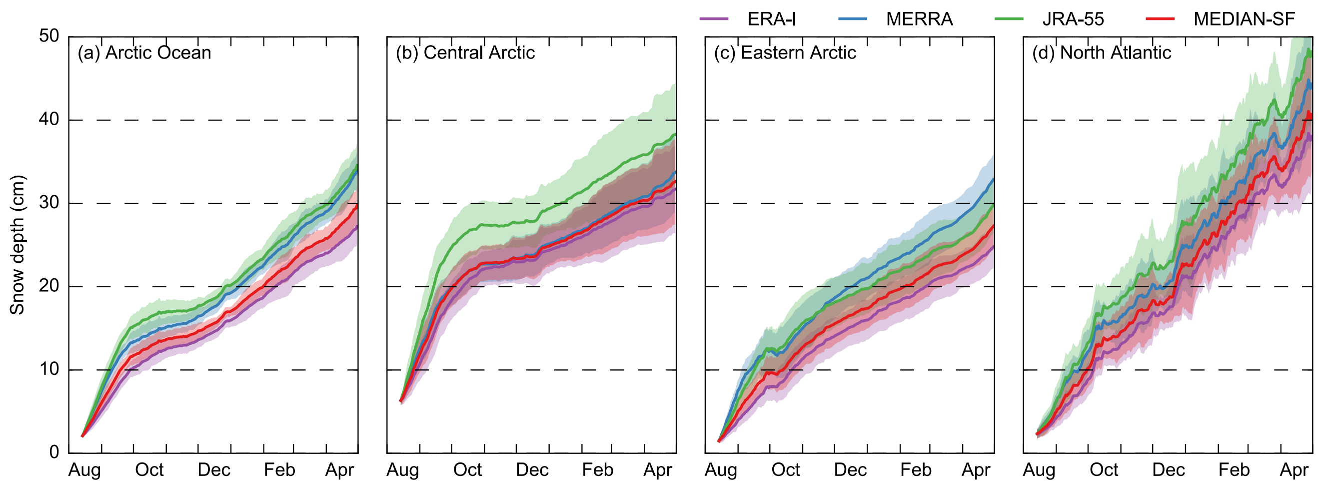


~~**Figure 6**: Seasonal snow depth (black) and bulk density (green) evolution across the four study regions (shown in Figure 3), initiated from August 15th 1980-1990 and run until May 1st of the following year using the MEDIAN-SF/parameter settings (Table 1). The thick lines show the mean values over this time period, while the shaded areas represent the interannual variability (one standard deviation).~~

~~We see stronger increases in the bulk snow density through fall across all regions (also shown in Figure 4), with this density increase slowing through winter/spring, especially in the CA region, after December. The AO, CA and NA regions also show an interesting initial decrease in snow density, which is driven by the accumulation of new snow (with a lower density) compared to the equal mix of old and new snow densities included in our initial conditions. The mean bulk snow densities as of May 1st are given as: 312 +/- 2 kg/m^3 (AO), 326 +/- 4 kg/m^3 (CA), 314 +/- 6 kg/m^3 (EA), 318 +/- 3 kg/m^3 (NA).~~

~~In Figure 7 we show the seasonal/regional snow depths from NESOSIM forced by the four different reanalysis-derived snowfall estimates (ERA-I, JRA-55, MERRA and MEDIAN-SF), as described in Section 3.1. In general, the results show significant differences in the seasonal snow depths across all regions (up to ~10 cm across all regions). The rankings of snow depth between the different products is broadly consistent across the four regions, with JRA-55 and MERRA producing consistently higher snow depths (except in the EA region where MERRA produces slightly higher snow depths), and ERA-I consistently lower. The MEDIAN-SF snow depths are, in~~

~~general, slightly higher than the ERA-I forced snow depths. In the CA region we can see that MERRA, ERA-I and MEDIAN-SF forced results are all broadly similar, with JRA-55 significantly higher (by ~5 cm from October onwards). It is thus expected that the MEDIAN-SF snowfall data will have excluded much of the high JRA-55 snowfall data (the benefits of using a median instead of a mean snowfall). Despite the NA region having the highest snow depths and interannual variability, the intra-reanalysis spread is similar to the other regions. The results further allude to the MEDIAN-SF dataset being a useful tool for producing estimates of snow depth considering the large uncertainty in reanalysis-derived snowfall. We further analyze the reanalysis sensitivity in the following section.~~

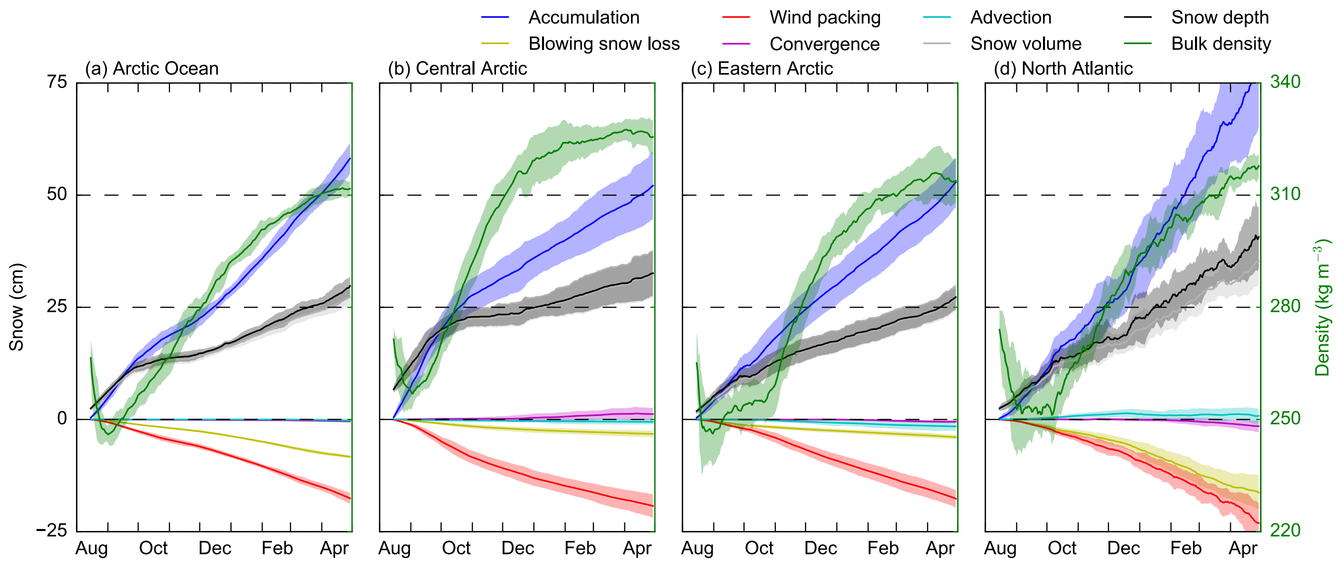


~~**Figure 7:** Seasonal snow depth evolution across the four study regions (shown in Figure 7), initiated from August 15th 1980-1990 and run until May 1st of the following year, forced by four different reanalysis snowfall products. The thick lines show the mean (daily) regional snow depths over this time period, while the shaded areas represent interannual variability (one standard deviation). All model runs use the default forcings/parameter settings.~~

~~4.2 Budget analysis~~

~~Here we discuss the relative contributions to the seasonal snow depth evolution from the various snow budget terms currently included in NESOSIM, focusing on the old time period results presented thus far. Results of the various NESOSIM budget terms and the total snow depth/volume and bulk density are shown in Figure 8 across our four study regions. The black (green) lines/shading that represent the snow depth (bulk density) are the same as the results shown in Figure 6.~~

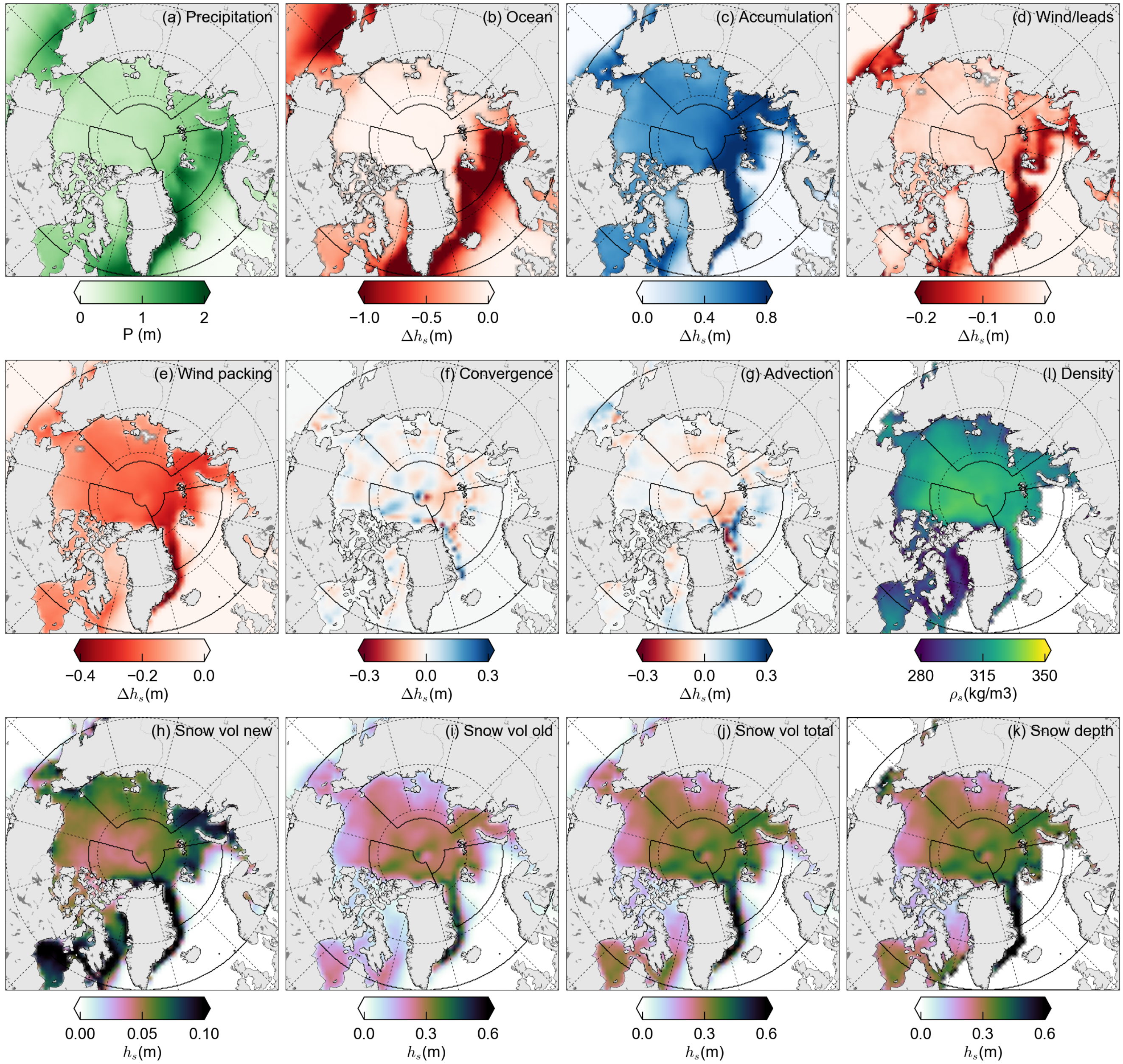
~~Across the AO region, we see that accumulation is higher than snow depth, as expected (higher by ~25 cm by May 1st, around double the May 1st snow depth), with wind packing (~18 cm) and wind blowing snow lost to leads (~9 cm), providing significant sinks of snow. In the EA and CA region especially, the blowing snow loss term is negligible, while in the NA region it is more significant (contributes a sink of ~18 cm by May 1st). The NA region also shows a small (~2 cm) increase in snow depth driven by snow/ice convergence and a seasonally variable change in snow depth from ice/snow advection.~~



~~**Figure 8**: Seasonal snow budget evolution across the four study regions (shown in Figure 5), initiated from August 15th 1980-1990 and run until May 1st of the following year using the default/MEDIAN_SF NESOSIM simulations. The thick lines show the mean, daily, regional values over this time period, while the shaded areas represent the interannual variability (one standard deviation).~~

~~To further explore the different budget terms we also show maps of the various budget terms as of May 1st over the same 1981-1991 time period, as shown in Figure 9. The maps highlight that many of these terms, especially the ice/snow dynamics (advection and convergence), exhibit high spatial variability, which the regional means discussed previously mask. For example, the NA region shows a strong mix of positive snow advection and convergence adjacent to the coast of Svalbard (i.e. snow is drifting into the region and is constrained against the coastline), but an advection out of the region further to the north as the ice either drifts down towards Svalbard/Fram Strait or into the Central Arctic.~~

~~The bimodal ice dynamic behaviour around the pole is thought to be spurious considering interpolating issues across the pole hole in the NSIDCv3 drift product (Szanyi et al, 2016), one reason why we did not include this region in our regional analysis. In the following section we assess the sensitivity of our results to the input ice drift dataset, which will provide some further information as to the reliability of these dynamic budget terms.~~



~~**Figure 9:** Snow budget terms as of May 1st, averaged over the 1981 to 1991 time period. The black lines show the four study regions used throughout this study. All model runs used the default forcings/parameter settings. Note the different color bar scales in panels (h) to (k).~~

44 Sensitivity studies and model validation

Here we present and analyze the NESOSIM results from 2000 to 2015, a period broadly defined as the New Arctic considering the rapid sea ice declines during this time period (e.g. Serreze and Stroeve, 2015). This period also covers the temporal range of NASA's ICESat (2003 to 2008) and ESA's CryoSat-2 (2010 onwards) satellite altimetry missions, meaning the snow depth/density results presented here are planned to be of more relevance ~~to~~ for those estimating sea ice thickness from these freeboard measurements. The period also includes temporal overlap with the ASR forcing data, and various satellite-derived ice ~~drift~~ motion products used. We provide examples of the model evaluation figures for the 1980s time period in the Supplementary Information (Figures S5 and S6). A more detailed study accounting for differences in the input forcing data is likely needed before any conclusions can be made regarding potential trends in seasonal Arctic snow depths, which is beyond the scope of this paper. We hope to explore trends in our simulated snow depths in future work, however.

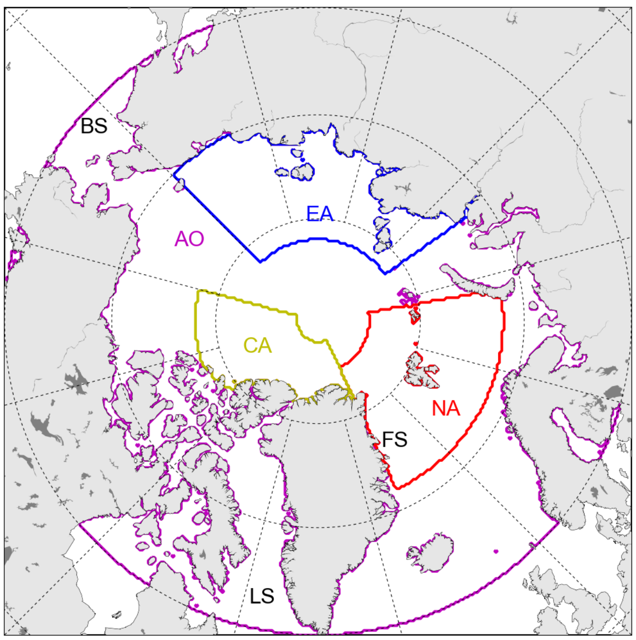


Figure 5: Map of the Arctic model domain and regions used in this study: AO: Arctic Ocean, CA: Central Arctic, EA: Eastern Arctic, NA: North Atlantic. BS: Bering Sea and LS: Labrador Sea are peripheral seas also discussed in the manuscript.

We focus our analysis on the Arctic Ocean (AO, everything north of 60 ^{o}N) and three specific regions that were chosen to represent different components of the Arctic sea ice/climate system: (i) the Central Arctic (CA, captures the thicker/multi-year ice over north of Greenland), (ii) the Eastern Arctic (EA, the increasingly first-year ice dominated sea ice regime), (iii) North Atlantic (NA, a region influenced by the transpolar ice drift and the North Atlantic storm track), as shown in Figure 5.

Figure 6 shows the seasonal snow depth and density evolution across our four study regions for the 2000-2015 time period, using the default/MEDIAN-SF configuration (Table 1). The AO and CA region especially show strong initial increases in snow depth through fall (August to October) with the snow depth increasing at a slower rate from November to May, which is in good agreement with the W99 climatology. The EA and NA regions show a more uniform increase in snow depth from August to April. The NA region shows more daily snow depth variability, which was expected due to the strong ice drifts and the location of the NA storm track where passing cyclones can deposit large quantities of snow in a short period of time. It is also worth noting the small decline in snow depth through September/October in the EA region which is driven by reduced snowfall and snow densification due to wind packing through this period. By May 1st the mean snow depths (and interannual variability, calculated as one standard deviation of the annual values) are given as: 27.8 +/- 1.9 cm (AO), 31.8 +/- 4.0 cm (CA), 23.2 +/- 2.9 cm (EA), 42.5 +/- 8.1 cm (NA). The May 1st snow depth results are summarized in Table 4, to aid comparison with the snow depths produced in the following sensitivity studies.

~~Results from NESOSIM forced with the MEDIAN-SF snowfall forcing and default settings from 2000 to 2015 are shown in Figure 10.~~

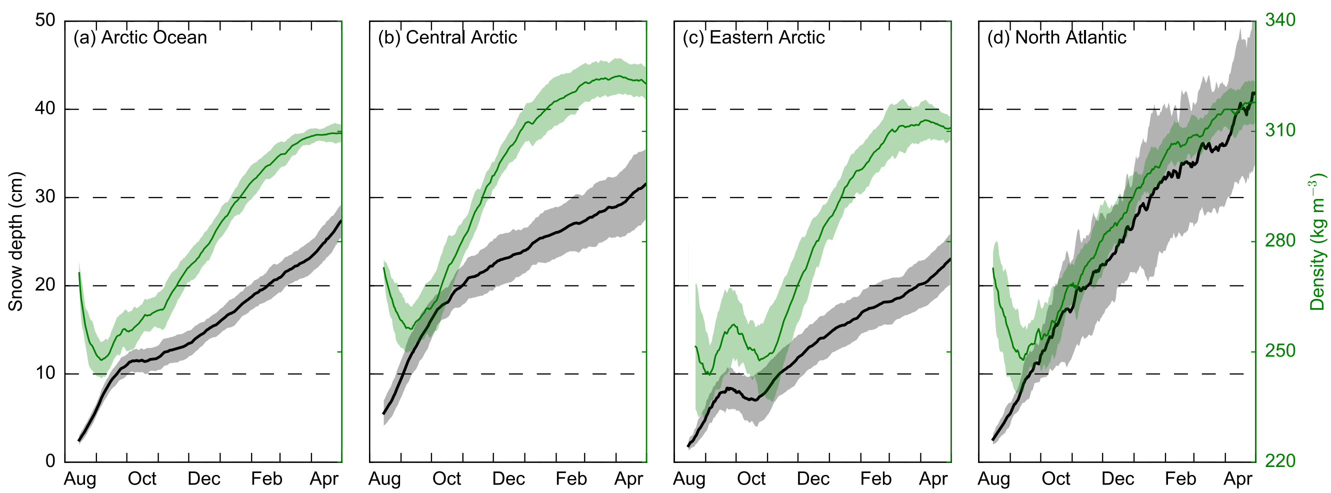


Figure ~~10~~6: Seasonal snow depth (black) and bulk density (green) evolution across the four study regions (shown in Figure 5) initiated from August 15th 2000-2014 and run until May 1st of the following year using the

MEDIAN-SF/parameter settings (Table 1). The thick lines show the mean values over this time period, while the shaded areas represent the interannual variability (one standard deviation). ~~As in Figure 6 but for the simulations initiated from August 15th 2000-2014 and run until May 1st of the following year. The simulations all use the default/MEDIAN-SF configuration.~~

~~The seasonal cycles over this New Arctic period are similar to the old period, but generally feature slightly reduced snow depths. The mean (2001-2015) May 1st snow depths are: 27.3 +/- 1.9 cm (AO), 31.5 +/- 4.0 cm (CA), 23.0 +/- 2.9 cm (EA), 41.8 +/- 8.1 cm (NA) while the mean May 1st bulk snow densities are: 309 +/- 2 kg/m^3 (AO), 323 +/- 4 kg/m^3 (CA), 311 +/- 3 kg/m^3 (EA), 318 +/- 6 kg/m^3 (NA). The May 1st snow depth results are summarized in Table 4, to aid comparison with the snow depths produced in the following sensitivity studies.~~ We see stronger increases in the bulk snow density through fall across all regions (also shown in Figure 4), with this density increase slowing through winter/spring, especially in the CA region, after December. The AO, CA and NA regions also show an interesting initial decrease in snow density, which is driven by the accumulation of new snow (with a lower density) compared to the equal mix of old and new snow densities included in our initial conditions. The mean bulk snow densities as of May 1st are given as: 309 +/- 2 kg/m^3 (AO), 323 +/- 4 kg/m^3 (CA), 311 +/- 3 kg/m^3 (EA), 318 +/- 6 kg/m^3 (NA).~~In the CA region, we see a more gradual increase in snow depth through fall compared to the old time period, while in the EA region we see an interesting decline in snow depth through September/October, which was not present in the old period except for the small period of constant snow depth at the start of October. While the NA region shows a similar snow depth as of May 1st between the two time periods, the new time period shows deeper snow depths in fall and winter. The results allude to strong regional variability in snow depth across the Arctic Ocean. A more detailed study accounting for differences in the input forcing data is likely needed before any conclusions can be made regarding potential trends in seasonal Arctic snow depths, which is beyond the scope of this paper. We hope to explore trends in our simulated snow depths in future work, however.~~

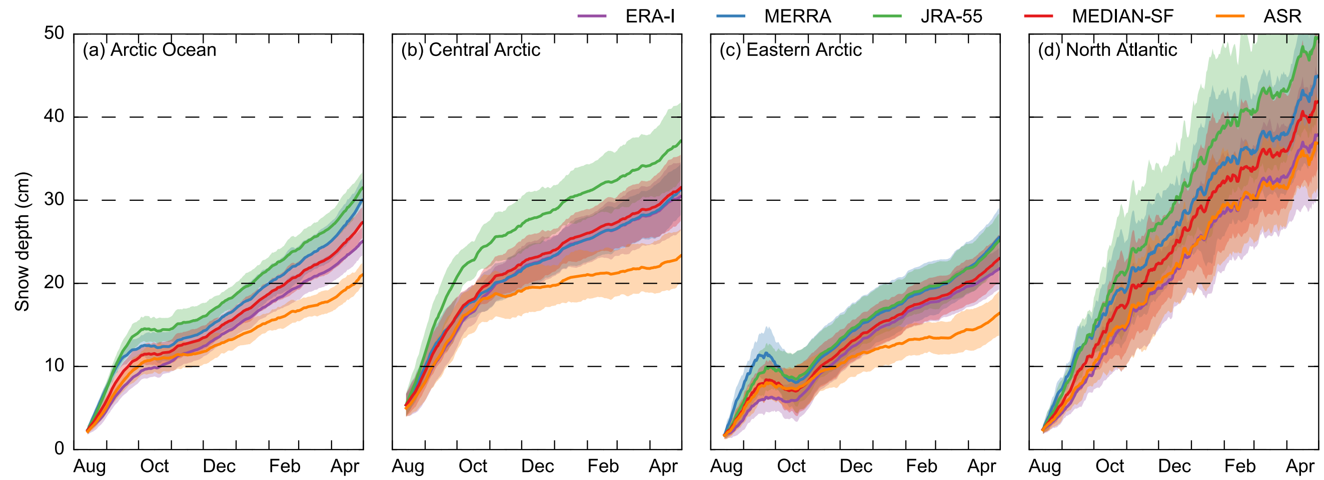
	May 1st snow depth (cm)			
NESOSIM configuration	**Arctic Ocean (AO)**	**Central Arctic (CA)**	**Eastern Arctic (EA)**	**North Atlantic (NA)**
Snowfall sensitivity results				
2001-2015 (ASRv1~~MEDIAN-SF~~)	21~~5.~~.4~~0~~ (1.5~~7~~)	23~~30~~.54	16.6~~21.7~~ (2.7~~5~~)	37.4~~9~~ (5~~7~~.4~~3~~)
2001-2015 (MERRA~~ERA-I~~)	30.6~~0~~ (2.6)	31.6~~3~~ (3.1)	25.8~~5~~ (3.6)	45~~4~~.7~~9~~ (9.0)
2001-2015 (JRA-55)	32~~1~~.0~~5~~ (1.9)	37.4~~1~~ (4.7)	25.4~~1~~ (3.3~~2~~)	50~~49~~.4~~5~~ (9.5)
2001-2015 (~~MERRA~~ERA-I)	25~~7~~.5~~3~~ (1.7~~9~~)	30~~1~~.7~~5~~ (4.1~~0~~)	22~~3~~.0 (2.5~~9~~)	348~~1~~.5~~8~~ (7.3~~8.1~~)

2001-2012 (~~ASRv1~~MEDIAN-SF)	27~~1~~.8~~0~~ (1.9~~4~~)	~~23.~~31.8	23~~16~~.24 (2.9~~7~~)	42~~36~~.59 (8.1~~(5.4~~)
Ice drift sensitivity results				
2011-2015* (MEDIAN-SF/NSIDCv3~~NODRIFT~~)	27~~6~~.3~~2~~ (2.2)	32.4~~2~~ (24.9~~6~~)	24~~2~~.8~~7~~ (4.0~~3.5~~)	38~~9~~.7~~2~~ (10.3~~9.0~~)
2011-2015* (MEDIAN-SF/OSISAF)	26~~5~~.3~~8~~ (2.2~~1~~)	32.9~~6~~ (3.8)	23.4~~2~~ (3.7~~6~~)	38.9~~1~~ (9.6~~5~~)
2011-2015* (MEDIAN-SF/KIMURA)	25.7~~3~~ (2.2)	32.9~~7~~ (4.3)	21.2~~0~~ (3.5)	38.9~~3~~ (10.6~~5~~)
2011-2015* (MEDIAN-SF/CERSAT)	26~~5~~.3~~9~~ (2.2~~1~~)	33~~2~~.1~~8~~ (3.9)	23.2~~0~~ (3.6)	38~~7~~.5~~7~~ (10.0~~9.9~~)
2011-2015* (MEDIAN-SF/~~NODRIFT~~NSIDCv3)	26.7~~8~~ (2.2)	32.4~~2~~ (4.6~~2.9~~)	22~~4~~.9~~6~~ (3.5~~9~~)	39~~3~~.9 (9.1~~10.2~~)
Ice concentration sensitivity results				
2001-2015 (MEDIAN-SF/NASA Team)	23~~2~~.4~~6~~ (1.7)	28~~7~~.0~~4~~ (4.0~~1~~)	20~~19~~.3~~8~~ (2.8~~7~~)	35~~4~~.4~~1~~ (7.8~~7~~)

Table 4: Mean snow depths as of May 1st across the four study regions (rows, regions given in Figure 5) for ~~different~~ NESOSIM using different forcings and time periods (columns). The numbers in brackets represent interannual variability and are calculated as one standard deviation of the annual values. The default NESOSIM configuration is MEDIAN-SF snowfall, ERA-I winds, NSIDCv3 ice motion and Bootstrap ice concentration *Note that these 2011-2015 ice ~~drift~~ motion sensitivity runs exclude the 2012-2013 winter season due to the lack of KIMURA drift data.

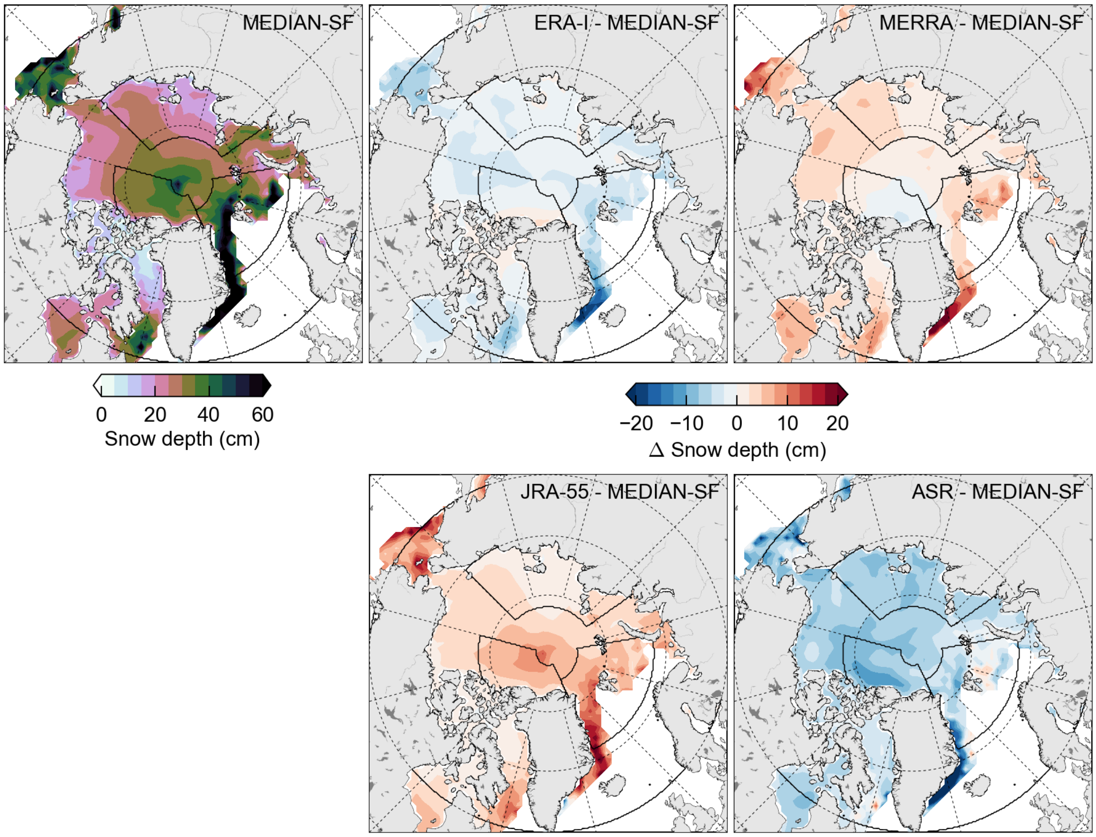
~~4.1 Reanalysis sensitivity~~

~~The various reanalysis forced May 1st results are summarized in Table 4. In Figure 11 we show the seasonal/regional snow depths from NESOSIM forced by the four different reanalysis-derived snowfall estimates over this new time period. We use the same reanalyses shown in the old time period sensitivity test (ERA-I, JRA-55, MERRA and MEDIAN-SF) but also include the ASRv1 forced results which are available during this period, but only up to 2012. The various reanalysis forced May 1st results are summarized in Table 4.~~



~~**Figure 11:** As in Figure 7 but for the simulations initiated from August 15th 2000-2014 and run until May 1st of the following year. This figure also includes results using the ASRv1 forced simulations (which are limited to Aug 15th 2000 to May 1st 2012).~~

~~In general, the results show similar sensitivity to the input snowfall data compared to the old time period results. The rankings of snow depth between the different products is also similar, except for the EA region, where MERRA now shows a clear high snow depth bias compared to the other forced simulations. The ASRv1 forced snow depths in the AO, CA and EA regions are significantly lower during the December-April time period, despite showing strong similarities to the other reanalysis-forced results in August to November. The ASRv1 results in the NA region however, are very similar to the ERA-I forced results. Note that we tested the impact of the different time periods by producing the same figure for the 2000-2012 period (not shown) which showed that the differences between ASRv1 and the other products was similar and not sensitive to this time period difference.~~



~~**Figure 12**: Modeled snow depth on May 1st (averaged over May 1st 2001 to 2015), using the MEDIAN-SF snowfall forcing (top left) and then the difference to the simulations forced by the four different snowfall products (bottom and top right). The ASRv1 forced results are limited to May 1st 2012.~~

4.2 Budget analysis

Here we discuss the relative contributions to the seasonal snow depth evolution from the various snow budget terms currently included in NESOSIM. Results of the various NESOSIM budget terms and the total snow depth and bulk density are shown in Figure 7 across our four study regions for this 2000s time period. The black (green) lines/shading that represent the snow depth (bulk density) are the same as the results shown in Figure 6.

Across the AO region, we see that accumulation is higher than snow depth, as expected (higher by ~30 cm by May 1st, around double the May 1st snow depth), with wind packing (~20 cm) and wind blowing snow lost to leads (~10 cm), providing significant reductions in snow depth. In the EA and CA region especially, the blowing snow loss term is negligible, while in the NA region it is more significant (contributes a sink of ~18 cm by May

1st). The NA region also shows a small (~2 cm) increase (decrease) in snow depth driven by snow/ice divergence (ice/snow advection).

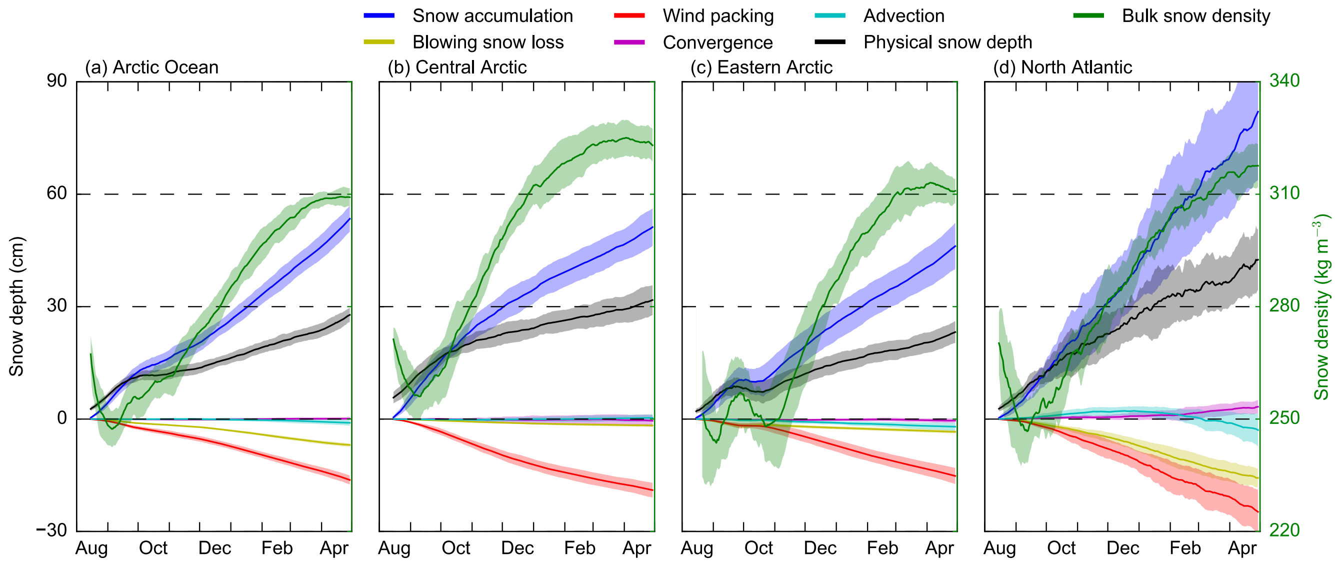


Figure 7: Seasonal snow budget evolution across the four study regions (shown in Figure 5), initiated from August 15th 2000-2014 and run until May 1st of the following year using the default/MEDIAN-SF NESOSIM simulations. The thick lines show the mean, daily, regional values over this time period, while the shaded areas represent the interannual variability (one standard deviation).

To further explore the different budget terms we also show maps of the various budget terms as of May 1st over the same time period, as shown in Figure 8. The maps highlight that many of these terms, especially the ice/snow dynamics (advection and convergence), exhibit high spatial variability, which the regional means discussed previously mask. For example, the NA region shows a strong mix of positive snow advection and convergence adjacent to the coast of Svalbard (i.e. snow is advected into the region and is constrained against the coastline), but an advection out of the region further to the north as the ice either drifts down towards Svalbard/Fram Strait or into the Central Arctic.

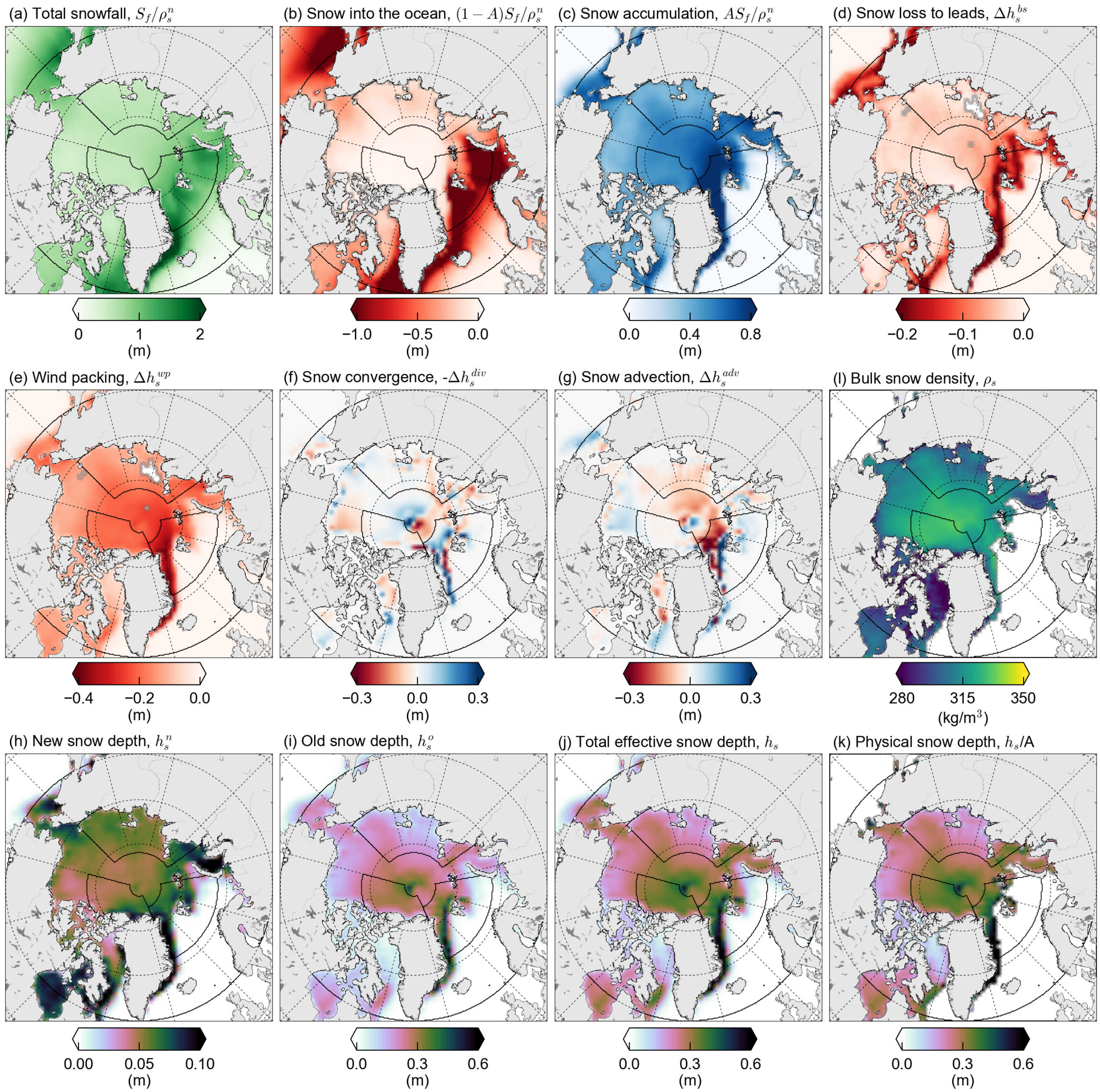


Figure 8: Snow budget terms as of May 1st, averaged over the 2001 to 2015 time period using the default/MEDIAN-SF NESOSIM simulations. The black lines show the four study regions used throughout this study. Note the different color bar scales in panels (h) to (k).

The ice dynamic behaviour around the pole is thought to be spurious considering interpolating issues across the pole hole in the NSIDCv3 drift product (Szanyi et al, 2016), one reason why we did not include this region in our regional analysis. In the following section we assess the sensitivity of our results to the input ice motion dataset, which will provide some further information as to the reliability of these dynamic budget terms.

As stated earlier, we hope to explore these decadal and regional differences more in future work. However, it is worth noting that t~~Figure 12 shows maps of the mean snow depths on May 1st over the same 2001-2015 time period, for the model simulations forced by the MEDIAN-SF snowfall then the differences from this MEDIAN-SF simulation using the four individual snowfall products. The maps highlight the regional variability across the products, but consistency in the MERRA/JRA-55 (ASR) high (low) bias compared to MEDIAN-SF. The JRA-55 and MERRA forced results both show significantly higher (10-20 cm) snow depths through Bering Strait, the NA/Fram Strait region, and the southern Labrador Sea. The ERA-I results show slightly lower snow depths over most of the Arctic, small increases around the Canadian Archipelago, and larger decreases in the Fram and Bering Strait region, driven by the larger differences in these regions in the MERRA/JRA-55 forcings. The magnitude of the precipitation events in Fram Strait are highly variable, due to the active storm track and the resulting difficulties of producing reliable precipitation rates during these events. As discussed in the old time period section, it is challenging to determine from this study any particular reanalysis-derived snowfall dataset that might be more appropriate (or an obvious outlier) for producing accurate snow depth estimates across the Arctic. However the MEDIAN-SF forced results appear to provide a useful synthesis of the available snowfall data.~~

~~T~~he regional snow budget results and May 1st budget maps using the same default/MEDIAN-SF ~~simulations~~ configuration but run ~~are similar to the figures presented for the~~for the ~~old~~ 1980s time period ~~(Figures 8 and 9)~~ were similar to the 2000s time period results (~~and are thus provided in the~~ Supplementary Information ~~(~~Figure S4 and S5). The noteworthy differences in the budget terms include a ~~more~~ less significant increase in blowing snow lost to leads in the CA region and ~~increases~~ less ~~in~~ convergent driven snow depth increases in the new period, although accumulation and wind packing still dominate the budget terms for both periods. The NA ~~region~~ results also ~~includes~~ do not show the~~an~~ ~~interesting~~ advection-driven reduction in snow depth in March/April that was ~~not~~ present in ~~the~~ our 2000s ~~old time period~~ results.

4.3 Reanalysis sensitivity study

In Figure 9 we show the seasonal/regional snow depths from NESOSIM forced by the various reanalysis-derived snowfall estimates (ERA-I, JRA-55, MERRA and MEDIAN-SF) from 2000 to 2015 and the ASRv1 forced results which are only available up to 2012, as described in Section 3.1. The May 1st results are summarized in Table 4. In general, the results show significant differences in the seasonal snow depths across all regions (up to ~10 cm across all regions). The rankings of snow depth between the different products is broadly consistent across the four regions, with JRA-55 and MERRA producing consistently higher snow depths (except in the EA region where MERRA produces slightly higher snow depths), and ERA-I consistently lower. The MEDIAN-SF snow depths are, in general, slightly higher than the ERA-I forced snow depths. In the CA region we can see that MERRA, ERA-I and MEDIAN-SF forced results are all broadly similar, with JRA-55 significantly higher (by ~5 cm from October onwards). It is thus expected that the MEDIAN-SF snowfall data will have excluded much of the high JRA-55 snowfall data (the benefits of using a median instead of a mean snowfall). Despite the NA region having the highest snow depths and interannual variability, the intra-reanalysis spread is similar to the other regions. The ASRv1 forced snow depths in the AO, CA and EA regions are significantly lower during the December-April time period, despite showing strong similarities to the other reanalysis-forced results in August to November. The ASRv1 results in the NA region however, are very similar to the ERA-I forced results. Note that we tested the impact of the different time periods by producing the same figure for the 2000-2012 period (not shown) which showed that the differences between ASRv1 and the other products was similar and not sensitive to this time period difference. The results further allude to the need for a consensus (e.g. median) snowfall dataset to force the model with considering the large uncertainty in reanalysis-derived snowfall.

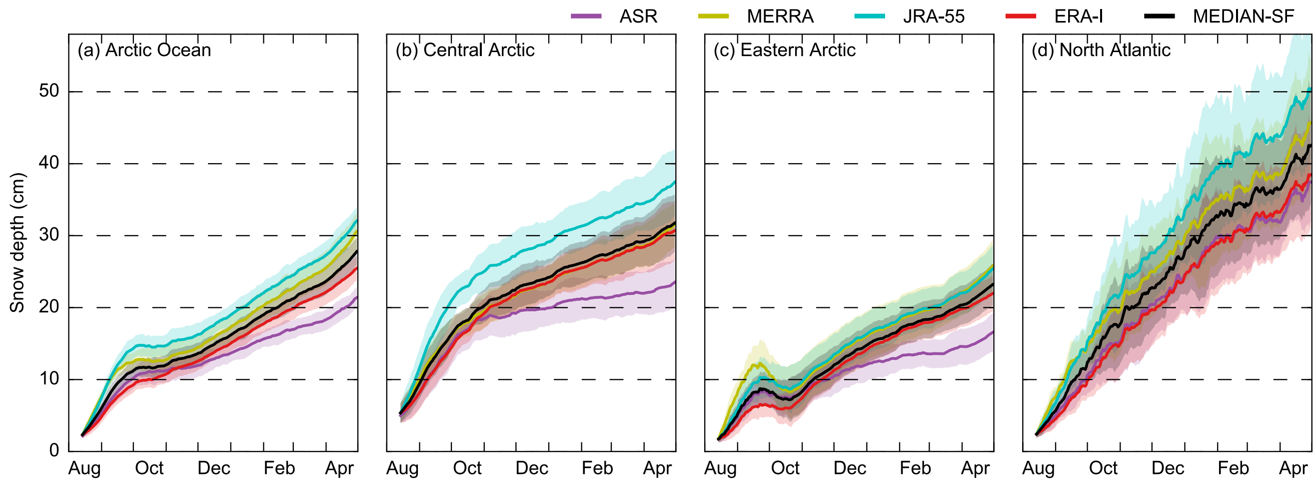


Figure 9: Seasonal snow depth evolution across the four study regions (shown in Figure 5) initiated from August 15th 2000-2014 and run until May 1st of the following year, forced by five different reanalysis snowfall products. This figure also includes results using the ASRv1 forced simulations (which are limited to Aug 15th 2000 to May 1st 2012). The thick lines show the mean (daily) regional snow depths over this time period, while the shaded areas represent interannual variability (one standard deviation). All model runs use the default forcings/parameter settings.

Figure 10 shows maps of the mean snow depths on May 1st over the same 2001-2015 time period, for the model simulations forced by the MEDIAN-SF snowfall then the differences from this MEDIAN-SF simulation using the four individual snowfall products.

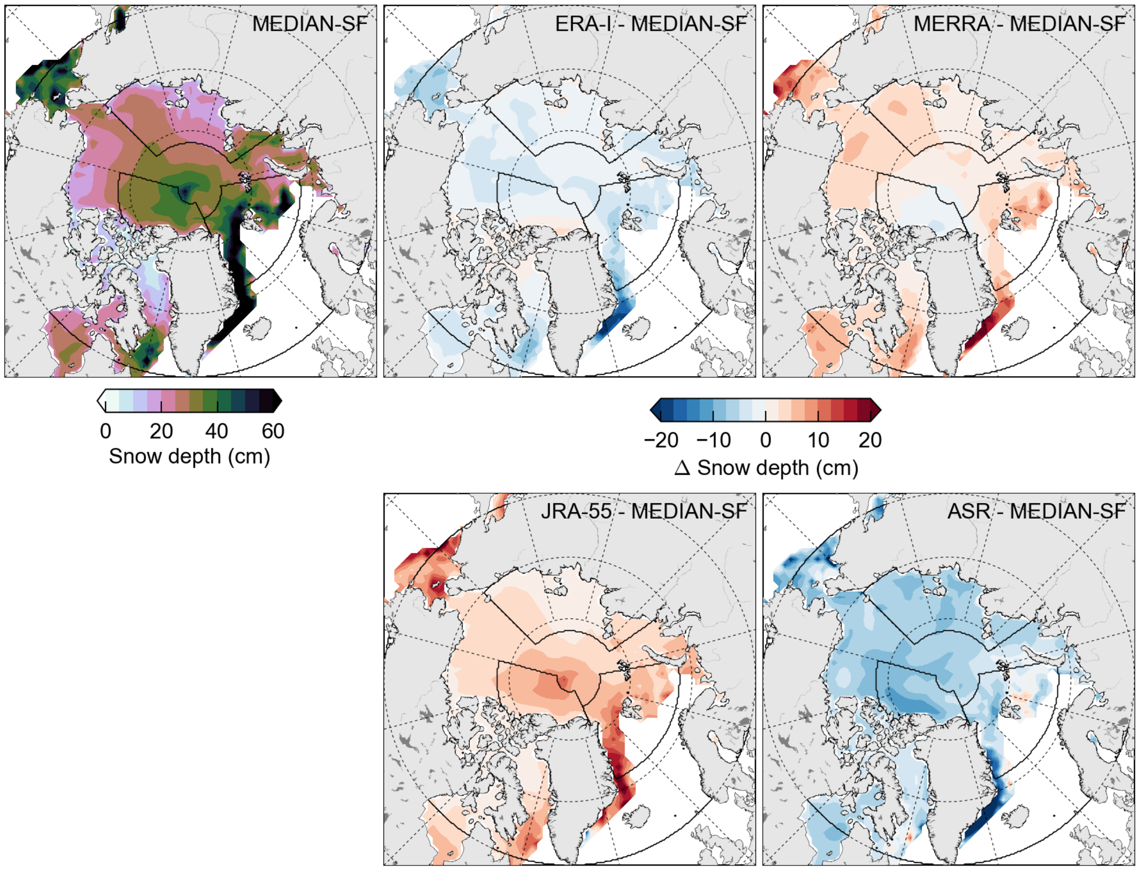
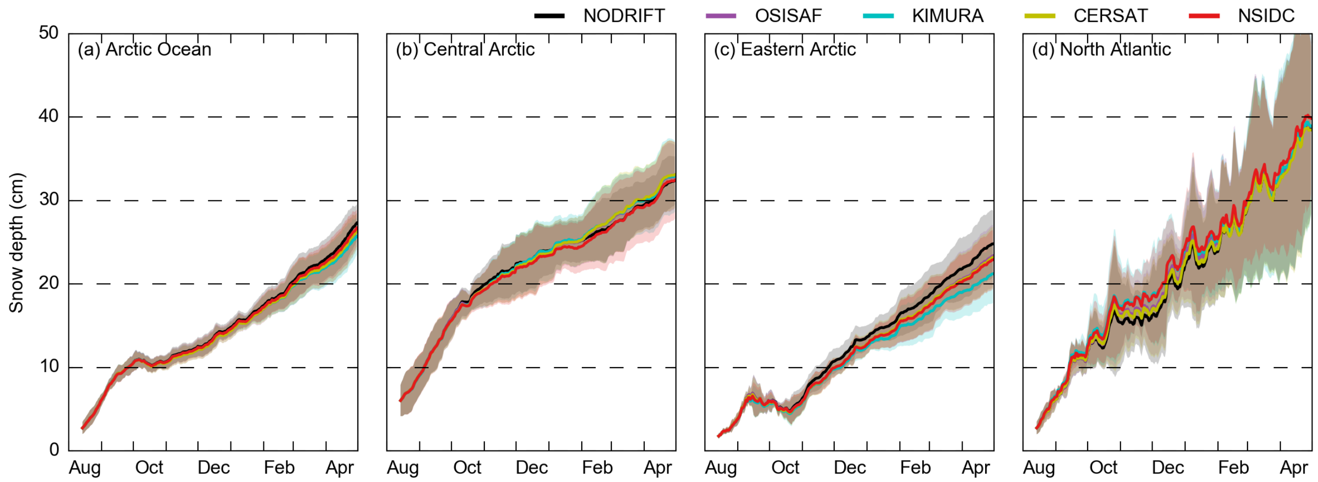


Figure 10: Simulated snow depths on May 1st (averaged over May 1st 2001 to 2015), using the MEDIAN-SF snowfall forcing (top left) and then the difference to the simulations forced by the four different snowfall products (bottom and top right). The ASRv1 forced results are limited to May 1st 2012. Red (blue) colours indicate the individual reanalysis-forced simulations have higher (lower) snow depth.

The maps highlight the regional variability across the products, but consistency in the MERRA/JRA-55 (ASR) high (low) difference compared to MEDIAN-SF. The JRA-55 and MERRA forced results both show significantly higher (10-20 cm) snow depths through Bering Strait, the NA/Fram Strait region, and the southern Labrador Sea. The ERA-I results show slightly lower snow depths over most of the Arctic, small increases around the Canadian Archipelago, and larger decreases in the Fram and Bering Strait region, driven by the larger differences in these regions in the MERRA/JRA-55 forcings. The magnitude of the precipitation events in Fram Strait are often large, but highly variable, due to the active storm track and the resulting difficulties of producing reliable precipitation rates during these events (Boisvert et al., 2018). As discussed earlier, it is challenging to determine from this study any particular reanalysis-derived snowfall dataset that might be more appropriate (or an obvious outlier) for producing accurate snow depth estimates across the Arctic. However, the MEDIAN-SF forced results appear to provide a useful synthesis of the available snowfall data.

4~~5~~.4~~2~~ Ice ~~drift~~ motion sensitivity

~~The newer time period allows us to~~We also explore the sensitivity of NESOSIM to the input satellite-derived ice ~~drift~~ motion data ~~due to the coincident data products~~ available during this period. Here we show results from the default/MEDIAN-SF configuration forced by four different satellite-derived ice drift products: NSIDCv3, KIMURA, CERSAT and OSISAF, as described in Section 3.2. Due to limitations in the temporal coverage of the different drift datasets, the model is only run for four years initialized from Aug 15th 2011-2015 (excluding 2012 initialized runs as KIMURA data are not available due to gaps in the AMSR-E/AMSR2 record). The regional snow depth estimates from NESOSIM forced by these four ice drift products are shown in Figure 11~~3~~, with the May 1st results summarized in Table 4. In general, the ice drift sensitivity study shows a smaller spread in the mean snow depths across the different products (up to ~3.5 cm), compared with the reanalysis sensitivity study (up to ~13~~4~~ cm). We also show results of NESOSIM forced with no ice drift (NODRIFT), which demonstrates that including ice drift appears not to be a crucial process for capturing the ~~regional~~ variability in snow depth at this regional scale~~s~~, i.e. ice dynamics appear to be a clear second order term compared to snowfall when ~~presented~~ analyzed at this regional scale. ~~The most obvious impact of ice drift is in the EA region, where the inclusion of ice drift reduces the snow depth by a few centimeters, with the magnitude depending on the ice drift product chosen (the KIMURA forced results shows the biggest decrease in this region).~~



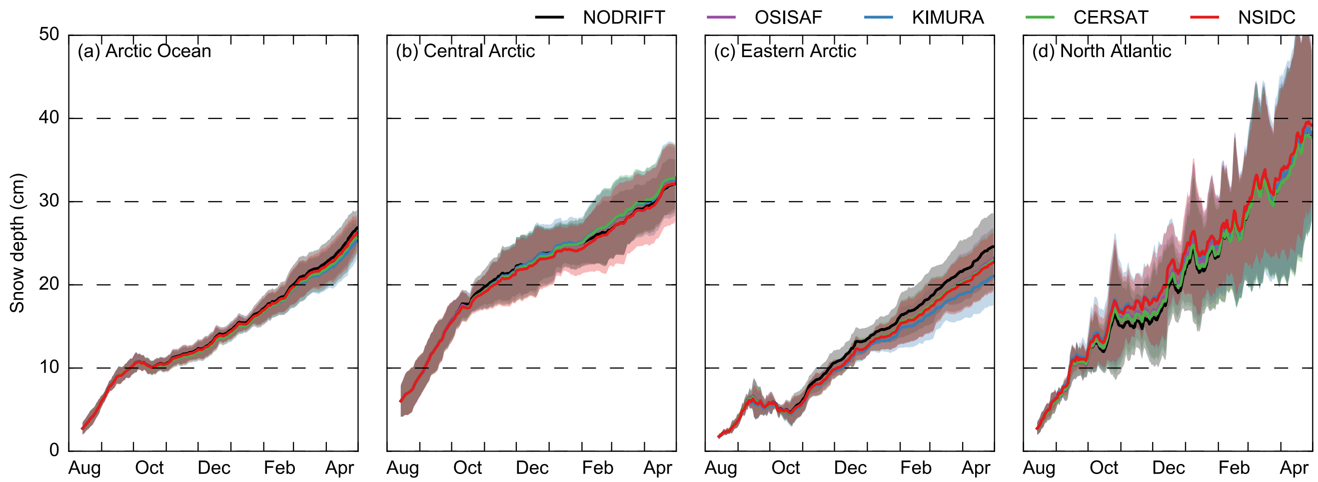
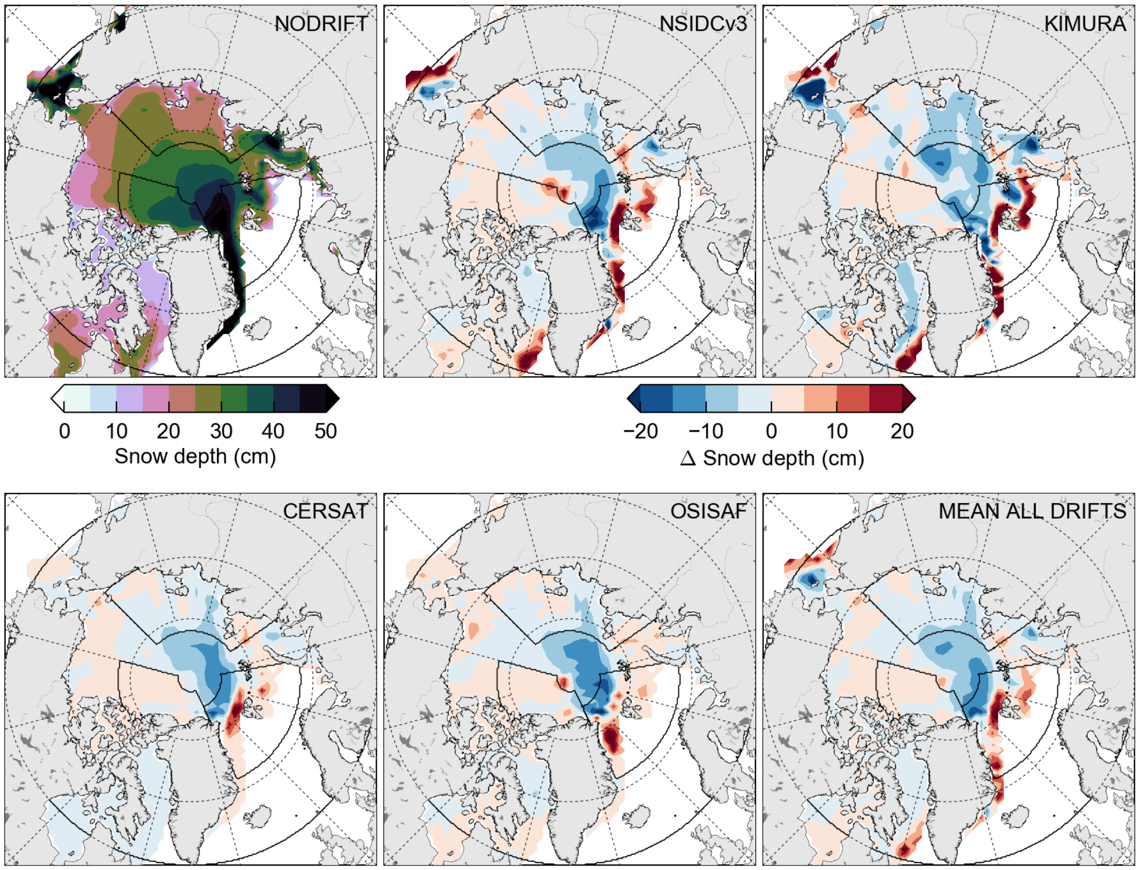


Figure 11~~13~~: Seasonal snow depth evolution across the four study regions (shown in Figure 5~~3~~), initiated in August 15th 2010, 2012, 2013, 2014 and run until May 1st of the following year, forced by four different ice ~~drift~~ motion datasets and assuming no ice ~~drift~~ motion (NODRIFT). The thick lines show the mean (daily) regional snow depths over this time period, while the shaded areas represent interannual variability (one standard deviation). All model runs use the default/MEDIAN-SF parameter settings.

The most obvious impact of ice drift is in the EA and NA regions. In the EA region the inclusion of ice drift reduces the snow depth by 1.4 - 3.6 cm, with the magnitude depending on the ice drift product chosen (the KIMURA forced results shows the biggest decrease in this region) In the NA region the inclusion of ice drift increases the snow depth by 3.8 to 5.3 cm (the NSIDCv3 forced results shows the biggest increase in this region).



~~Figure 14 shows maps of the snow depths averaged on May 1st over the same 2011-2015 time period, for the model simulations assuming no drift (NODRIFT) then the differences from this NODRIFT simulation using the various ice drift products. In general, the results show strong similarity in the spatial impacts of ice drift, including strong decreases in snow depth (up to ~10 cm) in the northeastern sector of the Arctic, and increases (up to ~10-20 cm) in the region directly north and west of Svalbard. There are clear differences between the different ice drift results though, with the NSIDCv3 and KIMURA forced results showing more of an impact on snow depth in the peripheral Arctic regions, e.g. strong decreases in the north and increases in the south Bering Strait, and strong increases in the Labrador Sea. This is thought to be driven primarily by the increased spatial coverage of these data compared to OSI-SAF and CERSAT, which may be masking some of the ice drift data in these regions of low ice concentration and uncertain ice drift. The maps also highlight that at more local scales, the ice dynamic contribution to snow depth variability could be significant. The data around the pole hole are also questionable in some of the products and may be related to interpolation issues across the pole hole. More specifically, the~~

~~NSIDCv3 and OSISAF forced simulations show increases in snow depth at the north pole, which are not apparent in the CERSAT and KIMURA simulations, suggesting this increase is likely spurious.~~

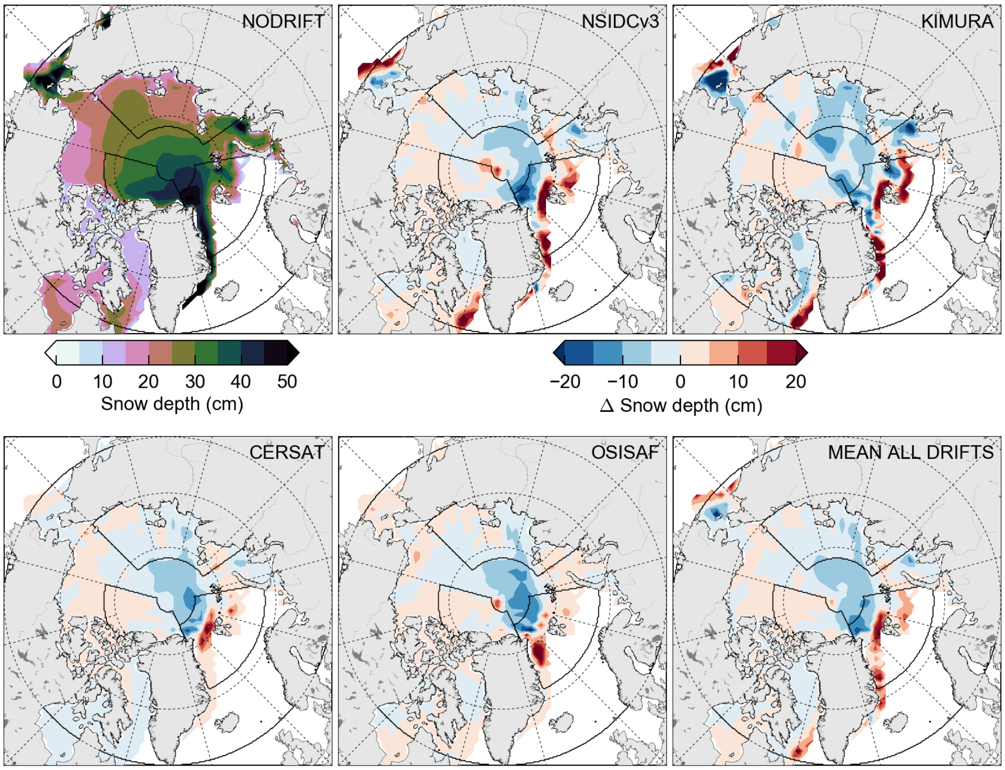


Figure 12~~4~~: Modeled snow depth on May 1st (averaged over May 1st 2011, 2013, 2014, 2015), assuming no ice ~~drift~~ motion (NODRIFT, top left) and then the difference to the simulations forced by the four different ice ~~drift~~ motion products and the mean snow depth from the four different forced model runs.

Figure 12 shows maps of the snow depths averaged on May 1st over the same 2011-2015 time period, for the model simulations assuming no drift (NODRIFT) then the differences from this NODRIFT simulation using the various ice motion products. In general, the results show strong similarity in the spatial impacts of ice motion, including strong decreases in snow depth (up to ~10 cm) in the northeastern sector of the Arctic, and increases (up to ~10-20 cm) in the region directly north and west of Svalbard. There are clear differences between the different ice motion results though, with the NSIDCv3 and KIMURA forced results showing more of an impact on snow depth in the peripheral Arctic regions, e.g. strong decreases in the north and increases in the south Bering Strait, and strong increases in the Labrador Sea. This is thought to be driven primarily by the increased spatial coverage of these data compared to OSI-SAF and CERSAT, which may be masking some of the ice motion data in these regions of low ice concentration and uncertain ice drift. The maps also highlight that at more local scales, the ice dynamic contribution to snow depth variability could be significant. The data around the pole hole are also questionable in some of the products and may be related to interpolation issues across the pole hole. More specifically, the NSIDCv3 and OSISAF forced simulations show increases in snow depth at the north pole, which are not apparent in the CERSAT and KIMURA simulations, suggesting this increase is likely spurious.

In general, Figures 11~~3~~ and 12~~4~~ suggest that the NSIDCv3 (Polar Pathfinder) forced simulations exhibit no obvious biases compared to the results using the other ~~drift~~ ice motion products, except for the issues of spurious snow depths within the pole hole and issues around the ice edge. ~~Note that another reason for exploring the ice drift products was to understand any potential biases if one of the near real-time products (e.g. OSI-SAF, CERSAT) were used to run NESOSIM in a near real-time framework, which does not appear to be the case.~~

4.5~~5.3~~ Ice concentration sensitivity

Finally we present and discuss the snow depth results from NESOSIM driven by two different satellite-derived ice concentration products (Bootstrap and NASA Team), as described in Section 3.3. The regional snow depth estimates from NESOSIM forced by these two ice concentration products over the ~~new~~ 2000-2015 time period are shown in Figure 13~~5~~, with the May 1st results summarized in Table 4. ~~In general, the ice concentration sensitivity study demonstrates that the choice of ice concentration product is significant, with differences of several centimeters between the two simulations across the study regions (e.g. > 7 cm differences in the NA snow depths).~~ In general, the ice concentration sensitivity study demonstrates that the choice of ice concentration product is significant, with differences of several centimeters between the two simulations across the study regions (e.g. ~~~>~~ 7 cm differences in the NA snow depths).

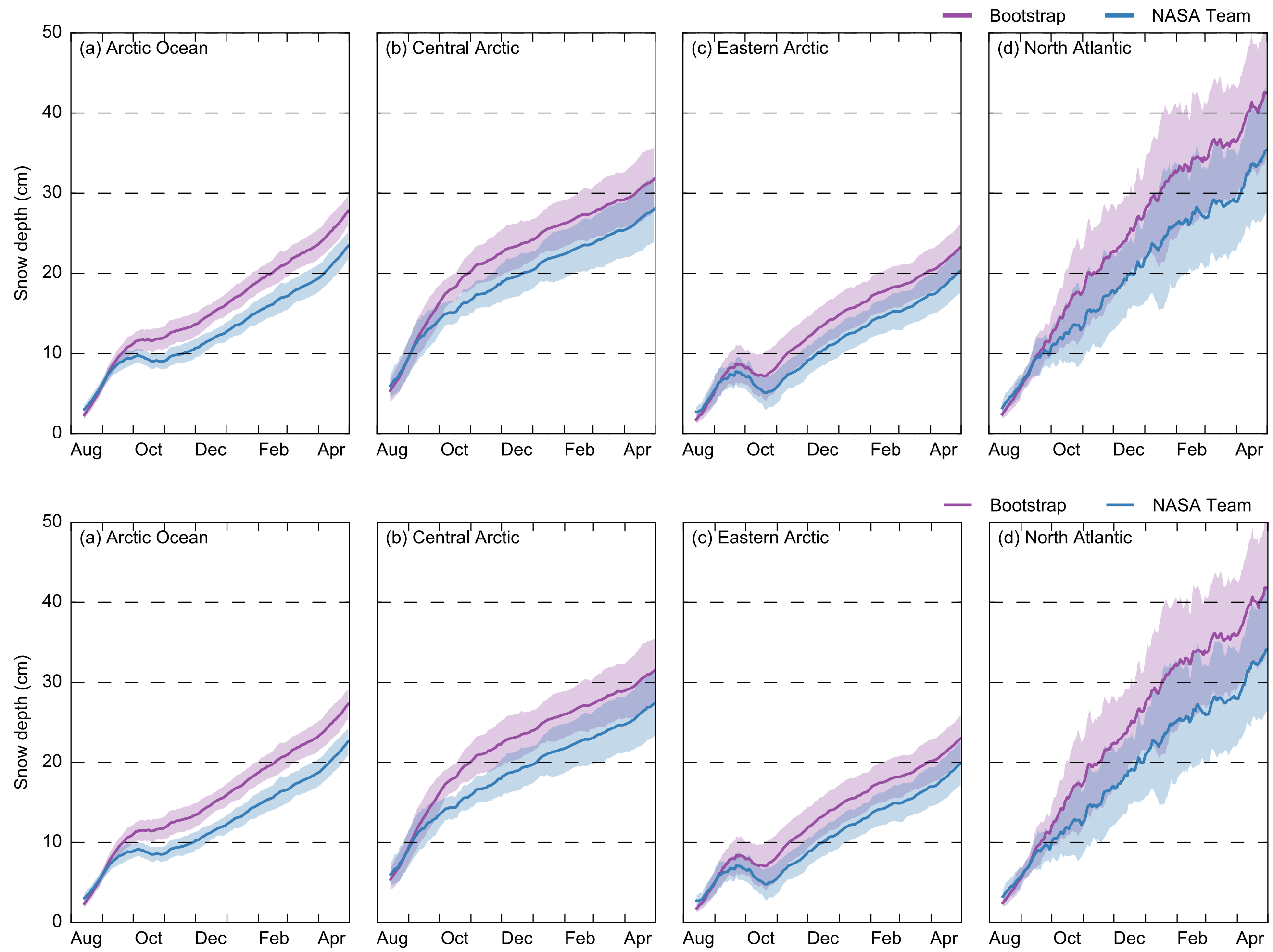


Figure 13~~5~~: Seasonal snow depth evolution across the four study regions (shown in Figure 5~~3~~), initiated from August 15th 2000-2014 and run until May 1st of the following year, forced by the Bootstrap (magenta) and NASA Team (blue) ice concentration datasets. The thick lines show the mean (daily) regional snow depths over this time period, while the shaded areas represent interannual variability (one standard deviation). All model runs use the default/MEDIAN-SF configuration.

This was somewhat expected given the known low bias in the NASA Team concentration data (e.g. Meier, 2005; Ivanova et al., 2015), reducing the concentration of sea ice for snow to accumulate on. More specifically the Bootstrap data use daily-variable tie-points and are thus thought to improve the distinction between surface melt and open water. The lower concentrations also increase the blowing snow lost to leads term (as this is a function of the open water fraction). The snow budget terms using the NASA Team concentration data are shown in the Supplementary Information (Figure S7~~6~~) to highlight this further, with all regions showing reduced snow accumulation and blowing snow lost to leads increased, and now significant, across all regions. Again, we believe the Bootstrap data better represent the seasonal ice conditions, although we appreciate uncertainties still remain regarding the treatment of surface melt/melt ponds and their impact on snow accumulation/depth.

4~~5~~.6~~4~~ Validation with Operation IceBridge data

Here we present and discuss comparisons of our NESOSIM snow depth estimates with NASA's Operation IceBridge spring snow depth data from 2009 to 2015, as described in Section 3.5. We first show the basin-averaged results for the various OIB snow depth products each spring (from 2009 to 2015) and the coincident NESOSIM snow depth estimates, to assess how well NESOSIM captures the mean snow depth and expected interannual snow depth variability across this broad region of the Arctic. As discussed in Section 3.5, the OIB flights mainly cover the western Arctic sea ice pack, broadly within and to the west of the Central Arctic domain used in our earlier regional analyses, although this does vary each year. Maps of the OIB snow depth results across the different products are given in Kwok et al., (2017).

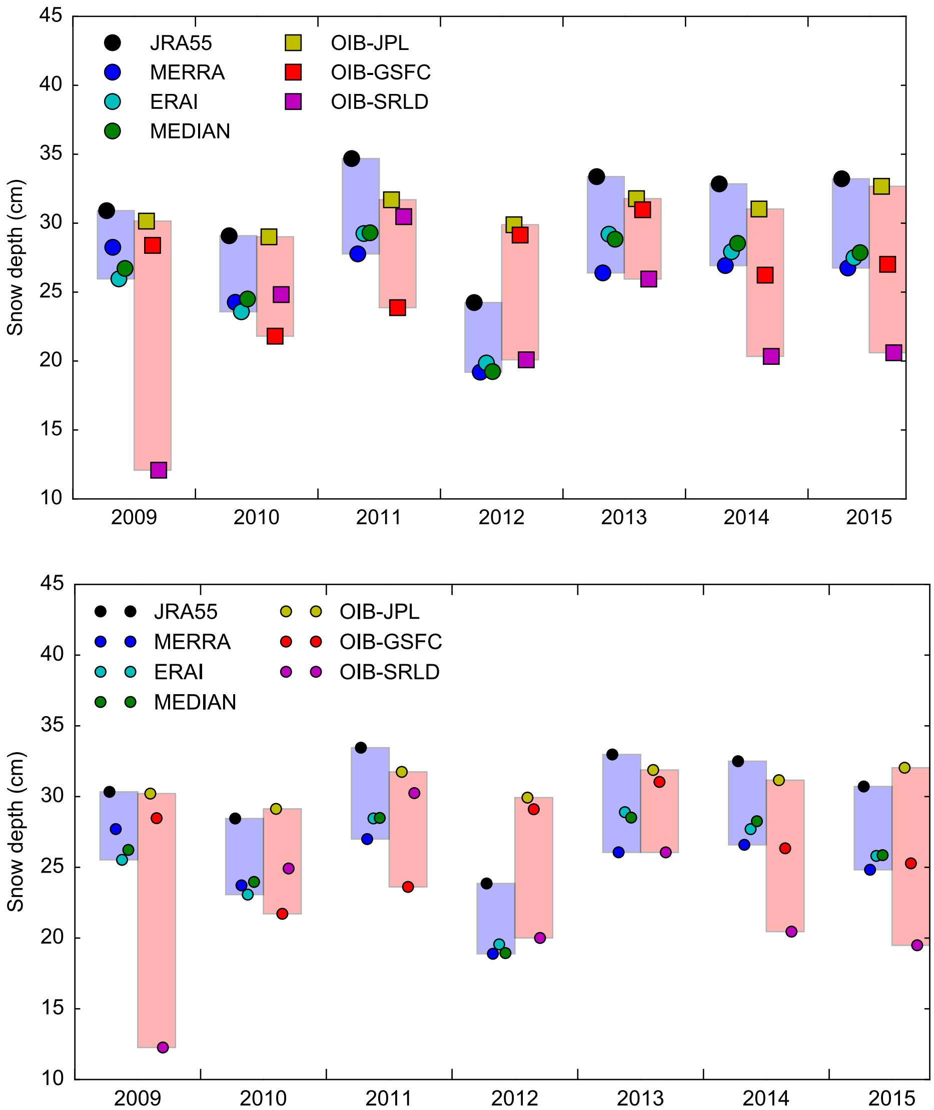
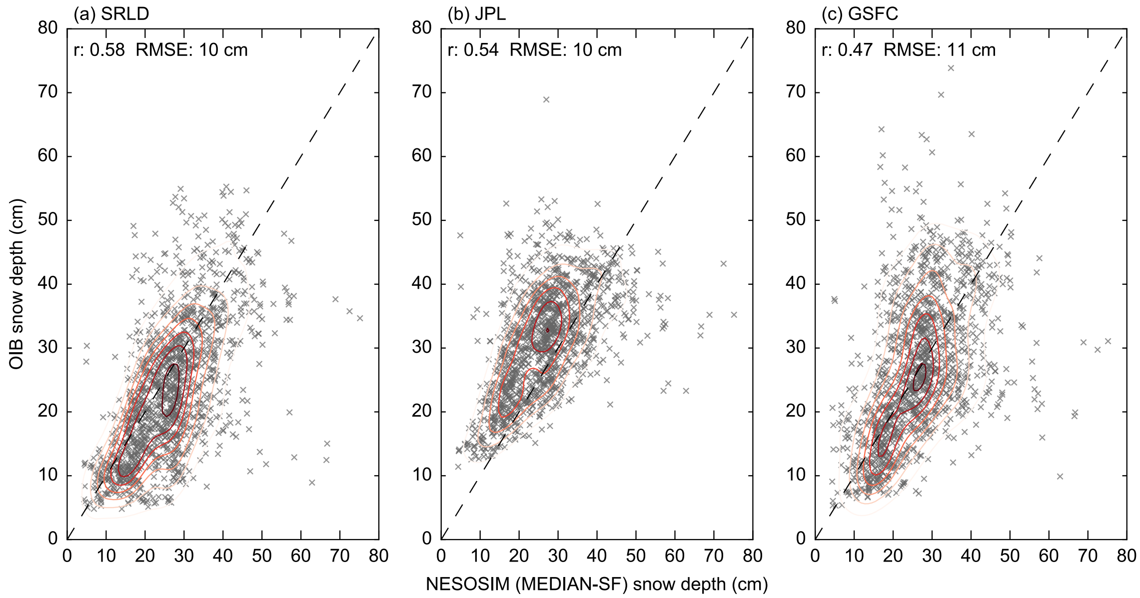


Figure 14~~16~~: Comparisons of the annual mean snow depths from NESOSIM (default configuration) forced with different reanalyses, and the various Operation IceBridge (OIB) snow depth products. The blue (red) shading represents the annual mean spread across the different NESOSIM results (OIB products). The markers are spread across the shaded areas to improve readability.

Figure 14~~6~~ highlights the significant and variable spread in the annual mean OIB snow depth estimates (product spread of ~5 to 20 cm depending on the year), with the OIB-JPL snow depths consistently higher and less variable than the other two OIB products (SRLD and GSFC). The reanalysis-forced NESOSIM snow depths exhibit a more consistent spread of ~5 cm, with the JRA-55 forced results consistently higher than the other reanalyses. This was expected based on our previous analyses (e.g. the Central Arctic results shown in Figure 11b). The large spread in the OIB snow depths make it challenging to assess the reliability and accuracy of our NESOSIM results. In general, however, there is broad agreement between the NESOSIM and OIB results in terms of the mean snow depths and the broad pattern of interannual variability.


(a) SRLD
r: 0.58 RMSE: 10 cm
(b) JPL
r: 0.54 RMSE: 10 cm
(c) GSFC
r: 0.47 RMSE: 11 cm
OIB snow depth (cm)
NESOSIM (MEDIAN-SF) snow depth (cm)
0
10
20
30
40
50
60
70
80

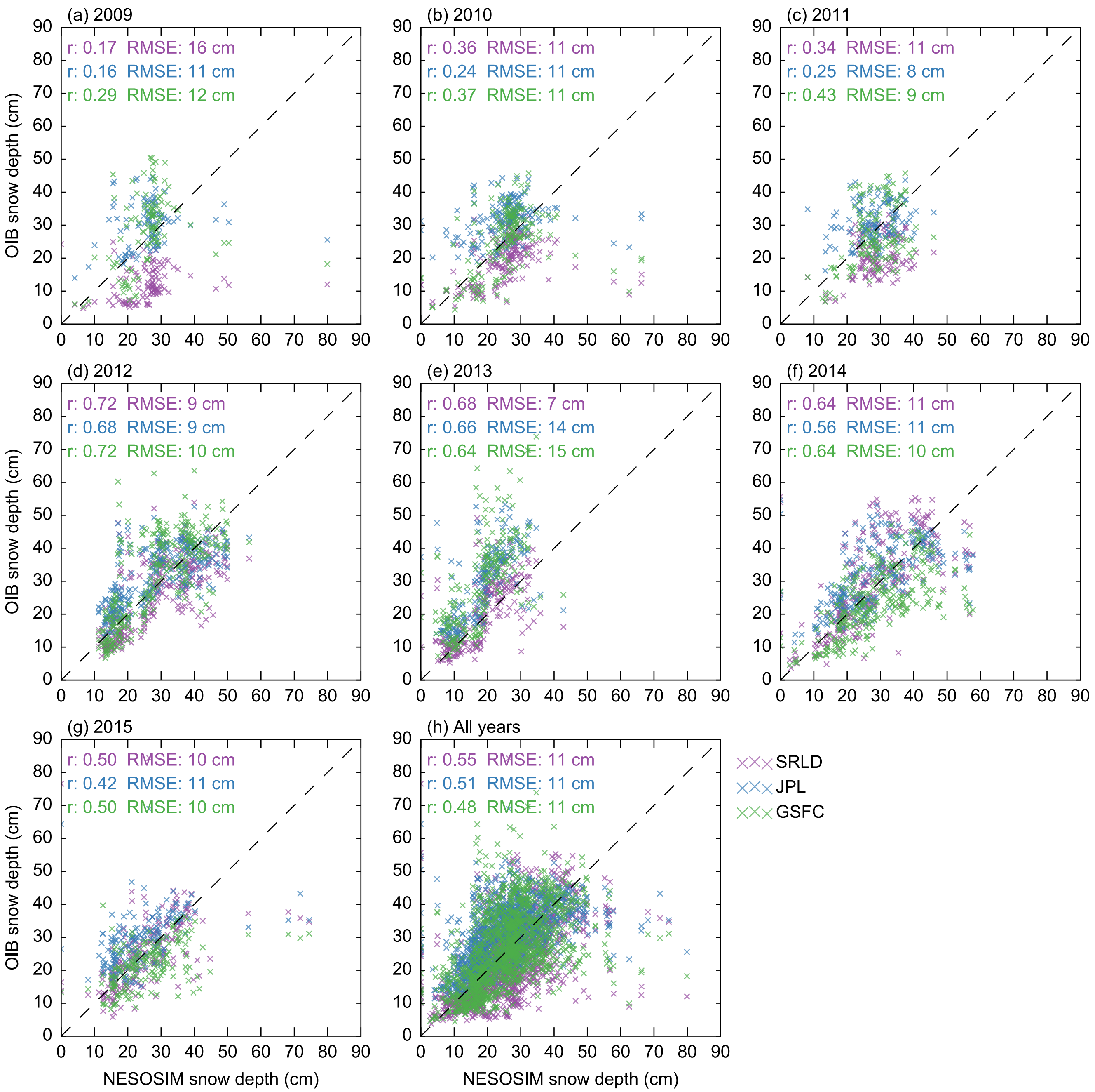


Figure 15~~7~~: Scatter plots of NASA's Operation IceBridge (OIB) snow depths from the three OIB products ~~interpolated~~ binned onto our 100 km model grid, and coincident NESOSIM/MEDIAN-SF snow depth estimates for ~~2009 to 2015 (a to g) and all years of data~~ all years of data from 2009-2015, including ~~(h), including~~ the correlation coefficient (r) and root mean squared error (RMSE). The contours show the kernel density estimate of the distributions.

To assess how well the model captures regional snow depth variability we show scatter plots in Figure 15~~7~~ of the MEDIAN-SF-NESOSIM snow depths and the three OIB snow depth products from 2009-2015 ~~.~~(the regressions for each year are given in Figure S8). A summary of the correlation coefficients (r) and root mean squared errors (RMSEs) across the three OIB products and NESOSIM forced by the three individual (and median) reanalysis products for individual years and for all years of data, are given in Table 5, with the regressions shown in Figures S9-11.~~.~~

In general, the comparisons are highly variable and depend mainly on the chosen analysis year and the reanalysis snowfall dataset, rather than the OIB product. The correlations between the OIB snow depths and the NESOSIM snow depths improve significantly in 2012 (r = 0.63~~1~~ - ~~to~~ 0.75~~4~~) compared to the proceeding years (r = -0.15~~01~~ - ~~to~~ 0.61~~59~~). The improved correlations in 2012 onwards coincide with increases in the OIB flight coverage, that include more of the Central Arctic and Beaufort/Chukchi seas, meaning the data better represent the regional variability in snow depths across the western Arctic. The strength of the correlations are highest in 2012 and 2013, while the RMSEs are lowest (< 10 cm) between 2011 and 2013, especially in the ERA-I and MEDIAN-SF forced results. The OIB-SLRD RMSEs are generally lower than the RMSEs calculated with the other OIB products between 2010 and 2015, but significantly higher in 2009 when the signal-to-noise ratio of the earlier version of the Snow Radar used on OIB was higher (Kwok et al., 2017). The 2009 OIB snow depth results should thus be treated with caution.

	MEDIAN-SF			**ERA-I**			**JRA-55**			**MERRA**		
Year	SRLD	JPL	GSFC	SRLD	JPL	GSFC	SRLD	JPL	GSFC	SRLD	JPL	GSFC
2009	0.27~~17~~ 17~~6~~ cm	0.17~~6~~ 11 cm	0.30~~29~~ 12 cm	0.37~~26~~ 15 cm	0.24~~3~~ 10 cm	0.36 11~~2~~ cm	0.32~~21~~ 21 cm	0.19~~8~~ 11 cm	0.36~~5~~ 12~~2~~ cm	0.18~~08~~ 19~~8~~ cm	0.12~~1~~ 11 cm	0.21~~0~~ 13 cm
2010	0.12~~36~~ 12~~1~~ cm	0.06~~24~~ 10~~1~~ cm	0.11~~37~~ 11 cm	0.27~~42~~ 11~~0~~ cm	0.27~~36~~ 9~~10~~ cm	0.29~~44~~ 10~~9~~ cm	0.16~~40~~ 16~~5~~ cm	0.07~~25~~ 11 cm	0.15~~40~~ 14~~3~~ cm	-0.06~~0.20~~ 13~~2~~	-0.15~~0.05~~ 12~~3~~	-0.08~~0.20~~ 13~~2~~

									cm	cm	cm
2011	0.38~~4~~ 11 cm	0.28~~5~~ 8 cm	0.46~~3~~ 9 cm	0.56~~3~~ 10 cm	0.47~~4~~ 7 cm	0.61~~59~~ 8 cm	0.29~~5~~ 16~~5~~ cm	0.20~~17~~ 9 cm	0.38~~5~~ 12 cm	0.14~~09~~ 10 cm	0.04~~–0.01~~ 9 cm	0.25~~3~~ 10 cm
2012	0.73~~2~~ 8~~9~~ cm	0.70~~68~~ 8~~9~~ cm	0.73~~2~~ 9~~10~~ cm	0.75~~3~~ 8 cm	0.72~~0~~ 8 cm	0.75~~4~~ 9 cm	0.72~~0~~ 12 cm	0.67~~6~~ 11 cm	0.72~~1~~ 10 cm	0.67~~5~~ 8 cm	0.63~~1~~ 9~~10~~ cm	0.66~~5~~ 11 cm
2013	0.69~~8~~ 7 cm	0.68~~6~~ 13~~4~~ cm	0.65~~4~~ 15 cm	0.73~~2~~ 7 cm	0.74~~2~~ 12~~3~~ cm	0.70~~68~~ 14 cm	0.67~~6~~ 9 cm	0.65~~3~~ 10~~1~~ cm	0.63~~2~~ 12~~3~~ cm	0.67~~6~~ 7~~8~~ cm	0.66~~4~~ 13~~4~~ cm	0.64~~3~~ 15 cm
2014	0.68~~4~~ 10~~1~~ cm	0.63~~56~~ 9~~11~~ cm	0.63~~4~~ 11~~0~~ cm	0.74~~69~~ 9~~10~~ cm	0.70~~63~~ 9~~10~~ cm	0.68 10 cm	0.64~~1~~ 12~~3~~ cm	0.58~~4~~ 12~~3~~ cm	0.60~~2~~ 15 cm	0.53~~0~~ 11~~2~~ cm	0.48~~2~~ 10~~2~~ cm	0.52~~3~~ 10 cm
2015	0.659~~50~~ 9~~10~~ cm	0.5~~4~~2 10~~1~~ cm	0.48~~50~~ 10 cm	0.6~~5~~8 8~~10~~ cm	0.62~~52~~ 9~~11~~ cm	0.54~~55~~ 9 cm	0.58~~49~~ 13~~3~~ cm	0.50~~41~~ 11~~2~~ cm	0.49~~50~~ 15 cm	0.37~~0~~ 10~~1~~ cm	0.29~~21~~ 11~~2~~ cm	0.32~~5~~ 10~~9~~ cm
All years	0.58~~5~~ 10~~1~~ cm	0.54~~1~~ 10~~1~~ cm	0.47~~8~~ 11 cm	0.64~~1~~ 9~~10~~ cm	0.62~~58~~ 9~~10~~ cm	0.53~~4~~ 10 cm	0.57~~5~~ 14~~4~~ cm	0.53~~0~~ 11 cm	0.47~~9~~ 13 cm	0.43~~2~~ 11 cm	0.41~~38~~ 11~~2~~ cm	0.37~~9~~ 12 cm

Table 5: Correlation coefficient (r, top rows) and root mean squared error (RMSE, bottom rows) from the correlations between the various reanalysis-forced NESOSIM results, and OIB derived snow depths. The MEDIAN-SF scatter plots for all years of data are shown in Figure 15~~7~~, with other reanalysis forced scatter plots given in the Supplementary Information.~~.~~

~~In general, the comparisons are highly variable and depend mainly on the chosen analysis year and the reanalysis snowfall dataset, rather than the OIB product. The correlations between the OIB snow depths and the NESOSIM snow depths improve significantly in 2012 (r = 0.61 to 0.74) compared to the proceeding years (r = -0.01 to 0.59). The improved correlations in 2012 onwards coincide with increases in the OIB flight coverage, that~~

~~include more of the Central Arctic and Beaufort/Chukchi seas, meaning the data better represent the regional variability in snow depths across the western Arctic. The strength of the correlations are highest in 2012 and 2013, while the RMSEs are lowest (< 10 cm) between 2011 and 2013, especially in the ERA-I and MEDIAN-SF forced results. The OIB-SLRD RMSEs are generally lower than the RMSEs calculated with the other OIB products between 2010 and 2015, but significantly higher in 2009 when the signal-to-noise ratio of the earlier version of the Snow Radar used on OIB was higher (Kwok et al., 2017). The 2009 OIB snow depth results should thus be treated with caution.~~

The 'all years' results in Table 5 provide a summary of the correlations using all the OIB snow depths from 2009-2015. The MERRA forced results produce significantly lower correlations to the OIB snow depths (r = 0.37~~8~~ to 0.43~~2~~) compared to the other reanalyses, while the ERA-I forced results show the highest correlations (r = 0.53~~8~~ to 0.64~~1~~) and lowest RMSEs (9 - 10 cm). The MEDIAN-SF results show slightly lower correlations (r = 0.47~~8~~ - ~~to~~ 0.58~~5~~) and higher RMSEs (10-11~~1~~ cm) compared to ERA-I. In general, however, the moderate to strong correlations give us confidence that NESOSIM is producing reasonable snow depth estimates across the western Arctic. The RMSEs of ~10 cm imply the expected level of accuracy in our NESOSIM snow depths, although these validations are hindered by uncertainty in the OIB snow depth observations (Kwok et al., 2017) and a lack of OIB observations in the eastern Arctic Ocean.

In the sensitivity studies presented earlier, we focused primarily on the MEDIAN-SF simulations, due in-part to considerations of high snowfall variability in regions of high and uncertain precipitation - e.g. the North Atlantic sector. The OIB data lack coverage in this region, however, making it hard to assess if this synthesized forcing snowfall produces more accurate snow depths in these more challenging regions of the Arctic. Data from the 2017 OIB flights into the eastern Arctic will hopefully provide some assessment of our NESOSIM snow depths in this region of the Arctic, however (the data was not available for this study but was made available during the review phase of this paper). Our contemporary (2000-2015) NESOSIM results still lack validation of the simulated snow densities, due to the lack of basin-scale density data available during this time period.

We can further assess the performance of NESOSIM by comparing these results with comparisons of OIB and the commonly used Warren snow depth climatology (Warren et al., 1999). As stated in the introduction, more recent uses of this climatology tend to apply a scaling factor (usually 50%) to the snow depths over first-year ice. We follow the same approach here, using the EUMETSAT Ocean and Sea Ice Satellite Application Facility (OSI SAF, www.osi-saf.org) ice type product which is derived from a combination of passive microwave and scatterometry data at 10 km horizontal resolution (Breivik et al., 2012). We derive daily modified W99 snow depths (referred to herein as mW99) for the same 100 km bins used in Figures 14 and 15 (where we have OIB data), with the comparisons shown in Figure 16.

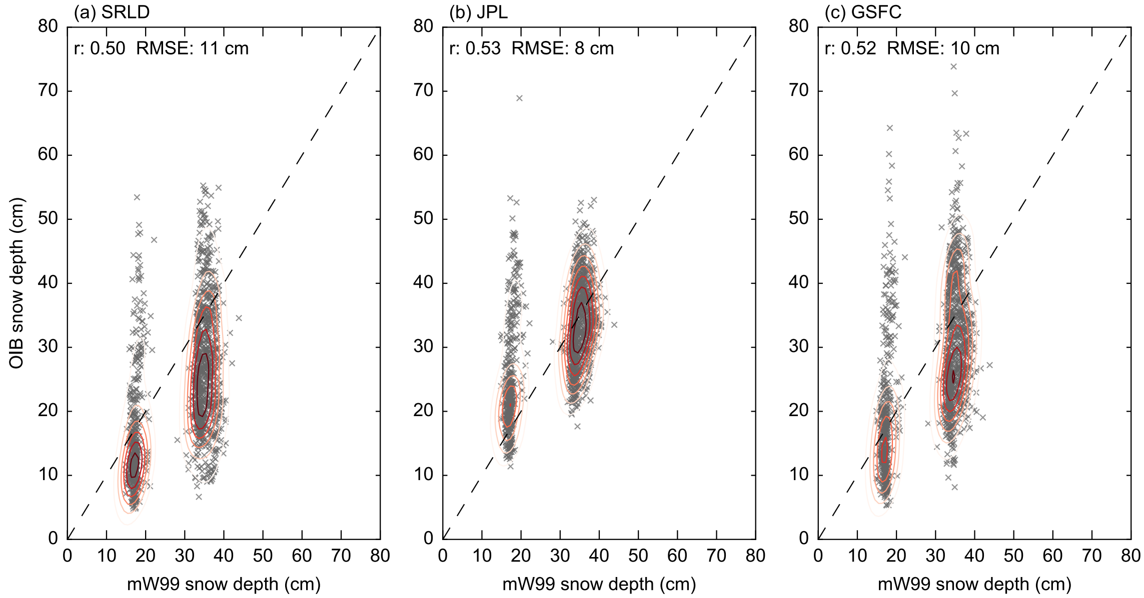


Figure 16: As in Figure 15 but showing comparisons of modified Warren snow depths (mW99) against the OIB snow depths.

In general these comparisons are similar, although in some cases the mW99 snow depths compare better with the OIB snow depth, depending on the product analyzed. Note that the bimodal nature of the mW99 data is due to the binary ice type weighting scheme, which does improve the comparison to the OIB data (comparisons of unmodified W99 data given in Figure S12). The low RMSE values in the JPL-OIB comparison is driven by the very good agreement in the mean snow depth, while the GSFC and SRLD products tend to show a slight low bias, increasing the RMSE values. Figure 14 shows that the OIB-JPL product exhibits less interannual variability than the other products, which may provide some explanation to the better correspondence with the climatology.

These results are useful for placing the NESOSIM comparisons with OIB snow depths in a wider context, which clearly show a higher spread and tend to suffer from positive and negative biases depending on the OIB product chosen. Bias correcting the NESOSIM snow depths could improve these comparisons, but uncertainty still remains regarding which OIB product better represents 'truth'.

~~In the sensitivity studies presented earlier, we focused primarily on the MEDIAN-SF simulations, due to considerations of snowfall reliability in regions of high and uncertain precipitation - e.g. the North Atlantic sector. The OIB data lack coverage in this region, however, making it hard to assess if this synthesized forcing snowfall produces more accurate snow depths in these more challenging regions of the Arctic. Data from the 2017 OIB flights into the eastern Arctic will hopefully provide some assessment of our NESOSIM snow depths in this region of the Arctic, however (the data has yet to be released). Our contemporary (New Arctic) NESOSIM results still lack validation of the simulated snow densities, due to the lack of basin-scale density data available during this time period.~~

6 Summary

In this study we presented the newly developed NASA Eulerian Snow On Sea Ice Model (NESOSIM). The snow depth and density simulated in NESOSIM (from August 15th to May 1st) across an Arctic Ocean domain (100 km horizontal grid) were first compared against in-situ data collected by drifting Soviet stations during the 1980s. The model produced very strong agreement with the seasonal cycles of snow depth and density and good (moderate) agreement with the regional snow depth (density) distribution. ~~A budget analysis provided insight into the relative processes contributing to the seasonal evolution in snow depth, with snow accumulation driving increases in snow depth, and wind packing reducing snow depth (through an increase in the bulk snow density). Blowing snow lost to leads provided a significant sink of snow, but only in the lower ice concentration, high wind/snow depth regime of the North Atlantic sector.~~

The model was run for a contemporary period (2000 to 2015) to produce seasonal snow depth and density estimates representative of the New Arctic climate system. A budget analysis provided insight into the relative processes contributing to our modelled seasonal evolution in snow depth, with snow accumulation driving increases in snow depth, and wind packing reducing snow depth (through an increase in the bulk snow density). Blowing snow lost to leads provided a significant sink of snow, but only in the lower ice concentration, high wind/snow depth regime of the North Atlantic sector.

The model showed strong sensitivity to the reanalysis-derived snowfall forcing data, with the MERRA/JRA-55 (ASR) derived snow depths generally higher (lower) than ERA-I. We derived a new synthesized snowfall dataset based on the median ERA-I, MERRA and JRA-55 snowfall data~~, to improve model reliability especially in regions of high/uncertain precipitation~~. We briefly assessed the sensitivity of NESOSIM to the input concentration data, with our results suggesting that the choice of concentration product (Bootstrap and NASA Team explored in this study) can have a significant impact on the simulations, and should not be overlooked. ~~The results across this newer period also allowed us to~~We also explored the sensitivity of NESOSIM to the input ice ~~drift~~ motion data, where we showed this had a second order effect compared to the choice of reanalysis snowfall forcing in our regional mean comparisons. The ice ~~drift~~ motion still appears to be important at smaller spatial scales, e.g. by reducing snow depths in the Eastern Arctic and driving higher snow depths north of Svalbard and within Fram Strait.

We compared our NESOSIM snow depths against spring snow depths derived from data collected by NASA's Operation IceBridge (OIB) since 2009 (up to spring of 2015) from three different algorithms. Our comparisons show moderate/strong correlations for the data collected from 2012-2015, but weaker correlations before this. The root mean squared differences were around 10 cm, but depend on the year analyzed, snowfall forcing and the OIB product analyzed. ~~with~~ T~~t~~he ERA-I and MEDIAN-SF forced results showing the best correspondence with the OIB snow depths. These results were compared with comparisons between OIB and the modified Warren snow depth climatology, which showed similar correlations and root mean squared errors.

~~These encouraging comparisons provide us with some confidence in our simulated daily NESOSOM snow depth and density estimates, however~~ W~~w~~e expect that further model development, ~~testing~~calibration, and validation is needed to improve accuracy and reliability in the NESOSIM snow depths/densities.

6.1 Future work

This initial formulation of NESOSIM (v1.0) has focused on: (i) incorporating several key snow parameterizations needed to capture the regional and seasonal variability in snow depth and density across the Arctic Ocean, and (ii) providing a framework simple and computationally efficient enough to run the various sensitivity studies needed to assess the importance of input forcing data. As highlighted throughout the manuscript, our relatively simple snow model is expected to undergo improvements to its model physics in efforts to increase its potential accuracy and reliability, together with further analyses of the input forcing data, especially snowfall (extending the precipitation comparison of Boisvert et al., 2018). Examples of expected future improvements to NESOSIM include the following, in order of priority:

- Incorporation of snow thermodynamics: modelling the temperature evolution of the snow pack and snow melt/refreeze processes, allowing us to run NESOSIM year-round. Challenges will include accurately modelling or parameterizing the temperature profile through the snow layers and the possible retention of meltwater within the snow pack and its impact on the snow density.

- Increased vertical snow layers: including depth hoar as an explicit snow layer as we introduce the snow thermodynamics described above. Model validation will be an obvious challenge.

- Snow-ice formation: NESOSIM is currently run independent of the sea ice state, meaning we include no information regarding the potential for snow-ice formation - the depression of the snow layer below sea level and the conversion of snow to ice. This is thought to be particularly important for running NESOSIM in the Southern Ocean (e.g. Massom et al., 2001), where snow-ice conversion is expected to be more prevalent, but also in our North Atlantic sector (e.g. Granskog et al., 2017). Challenges will involve incorporating observed or simulated sea ice freeboard.

- Increased horizontal resolution: As we look towards the launch of NASA's ICESat-2 and the production of sea ice thickness from the derived freeboard product, we hope to increase the model resolution and conduct assessments of the ability of NESOSIM to capture smaller-scale (< 100 km) snow depth variability.

~~As we look towards the launch of NASA's ICESat-2 and the production of sea ice thickness from the derived freeboard product, we must also consider potential increases in model resolution, and a better assessment of the ability of NESOSIM to capture smaller-scale (< 100 km) snow depth variability.~~ Snow depth and density information collected during the Norwegian young sea ICE (N-ICE2015) expedition (Merkouriadi et a., 2017) and the upcoming Multidisciplinary drifting Observatory for the Study of Arctic Climate (MOSAiC) will provide crucial insight into the importance of smaller-scale phenomena not currently included in NESOSIM, while our model results can hopefully provide useful basin-scale context to the measurements being taken.

NESOSIM is being made available as an open source project (https://github.com/akpetty/NESOSIM), to encourage continued model development and active engagement with the snow on sea ice community. The model code is written in Python, an open source programming language (Python Software Foundation, https://www.python.org/), to better enable future community development efforts. Our hope is that the model will continue to evolve as additional snow processes are incorporated, especially as new field and remote sensing snow observations are collected and made available for calibration/validation. ~~Obvious examples of planned future improvements include the incorporation of snow-ice formation, snow melt and rain on snow processes, which are not currently included in this initial model version, enabling the model to be run year-round.~~

~~As we look towards the launch of NASA's ICESat-2 and the production of sea ice thickness from the derived freeboard product, we must also consider potential increases in model resolution, and a better assessment of the ability of NESOSIM to capture smaller-scale (< 100 km) snow depth variability. Snow depth and density information collected during the Norwegian young sea ICE (N-ICE2015) expedition (Merkouriadi et a., 2017) and the upcoming Multidisciplinary drifting Observatory for the Study of Arctic Climate (MOSAiC) will provide crucial insight into the importance of smaller-scale phenomena not currently included in NESOSIM, while our model results can hopefully provide useful basin-scale context to the measurements being taken.~~

Model availability

All the data processing and figure generation was carried out using the Python programming language (Python Software Foundation, https://www.python.org/). The model code, including installation details and test data, can be found on GitHub (https://github.com/akpetty/NESOSIM).

Data availability

A link to the model output (hosted on the NASA Cryospheric Sciences website) will be made available after completion of peer review, along with the gridded OIB snow depths and KIMURA ice ~~drift~~ motion data.

The ERA-I snowfall and wind data were obtained through the ECWMF Meteorological Archival and Retrieval System (http://apps. ecmwf.int/datasets/data/interim_full_ daily/). The JRA-55 snowfall data were obtained through the NCEP Research Data Archive (RDA) (http://rda.ucar.edu/ datasets/ds628.0). The MERRA snowfall data were obtained through the NASA Goddard Earth Sciences Data and Information Services Center (https://disc.sci.gsfc.nasa.gov/datasets?page=1&keywords=merra).

The sea ice concentration data were obtained through the National Snow and Ice Data Center (NSIDC), including daily NASA Team (http://nsidc.org/data/nsidc-0051) and Bootstrap (https://nsidc.org/data/nsidc-0079) data.

The NSIDCv3 Polar Pathfinder ice ~~drift~~ motion data were obtained through the NSIDC (http://nsidc.org/data/nsidc- 0116). The CERSAT ice ~~drift~~ motion data were obtained from the IFREMER website (ftp://ftp.ifremer.fr/ifremer/ cersat/products/gridded/psi-drift/). The OSI-SAF ~~data~~ ice motion data were obtained through their web portal (http://osisaf.met.no/p/ice/).

Acknowledgements

All authors were funded for this work through the NASA-ESA Snow On Sea Ice (NESOSI) project.

References

Armitage, T. W. K. and Ridout, A. L.: Arctic sea ice freeboard from AltiKa and comparison with CryoSat-2 and Operation IceBridge, Geophys. Res. Lett., 42(16), 2015GL064823, doi:10.1002/2015GL064823, 2015.

Blanchard-Wrigglesworth, E., Webster, M. A., Farrell, S. L. and Bitz, C. M.: Reconstruction of Snow on Arctic Sea Ice, J. Geophys. Res. Oceans, 123(5), 3588–3602, doi:10.1002/2017JC013364, 2018.

Boisvert, L. N., Webster, M. A., Petty, A. A., Markus, T., Bromwich, D. H. and Cullather, R. I.: Intercomparison of Precipitation Estimates Over the Arctic Ocean and its peripheral seas from Reanalyses, J. Clim., doi:10.1175/JCLI-D-18-0125.1, 2018.

Breivik, L. A., Eastwood, S. and Lavergne, T.: Use of C-Band Scatterometer for Sea Ice Edge Identification, IEEE Trans. Geosci. Remote Sens., 50(7), 2669–2677, doi:10.1109/TGRS.2012.2188898, 2012.

~~Boisvert, L. N., M. A. Webster, A. A. Petty, T. Markus, D. H. Bromwich and R. I. Cullather, Intercomparison of precipitation estimates over the Arctic Ocean from Reanalyses. J. Climate (submitted).~~

~~Blanchard-Wigglesworth, E., M. Webster, S. L. Farrell, C. Bitz, Reconstruction of Arctic snow on sea ice, J. Geophys. Res, (accepted).~~

Bromwich, D. H., Wilson, A. B., Bai, L.-S., Moore, G. W. K. and Bauer, P.: A comparison of the regional Arctic System Reanalysis and the global ERA-Interim Reanalysis for the Arctic, Q.J.R. Meteorol. Soc., 142(695), 644–658, doi:10.1002/qj.2527, 2016.

Brucker, L. and Markus, T.: Arctic-scale assessment of satellite passive microwave-derived snow depth on sea ice using Operation IceBridge airborne data, J. Geophys. Res. Oceans, 118(6), 2892–2905, doi:10.1002/jgrc.20228, 2013.

Cavalieri, D., Parkinson, C., Gloersen, P. and Zwally, J. H.: Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS Passive Microwave Data, Version 1. [January 2000 - December 2010]. Boulder, Colorado USA: NASA DAAC at the National Snow and Ice Data Center, , doi:10.5067/8GQ8LZQVL0VL., 1996.

Colbeck, S. C.: An overview of seasonal snow metamorphism, Rev. Geophys., 20(1), 45–61, doi:10.1029/RG020i001p00045, 1982.

Comiso, J. C.: Bootstrap Sea Ice Concentrations from Nimbus-7 SMMR and DMSP SSM/I-SSMIS. Version 2. [January 2000 - December 2010]. Boulder, Colorado USA: NASA DAAC at the National Snow and Ice Data Center, , doi:10.5067/J6JQLS9EJ5HU., 2000.

Comiso, J. C., Cavalieri, D. J. and Markus, T.: Sea ice concentration, ice temperature, and snow depth using AMSR-E data, IEEE Transactions on Geoscience and Remote Sensing, 41(2), 243–252, doi:10.1109/TGRS.2002.808317, 2003.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Holm, E. V., Isaksen, L., Kallberg, P., Kahler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J.-N. and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q.J.R. Meteorol. Soc., 137(656), 553–597, doi:10.1002/qj.828, 2011.

Eicken, H., Krouse, H. R., Kadko, D. and Perovich, D. K.: Tracer studies of pathways and rates of meltwater transport through Arctic summer sea ice, J. Geophys. Res., 107(C10), SHE 22-1–SHE 22-20, 2002.

Eicken, H., Grenfell, T. C., Perovich, D. K., Richter-Menge, J. A. and Frey, K.: Hydraulic controls of summer Arctic pack ice albedo, J. Geophys. Res., 109(C8), C08007, doi:10.1029/2003JC001989, 2004.

Fetterer, F. and Radionov, V. F.: Environmental Working Group Arctic Meteorology and Climate Atlas, Version 1. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center, , doi:10.7265/N5MS3QNJ., 2000.

Giles, K. A., Laxon, S. W., Wingham, D. J., Wallis, D. W., Krabill, W. B., Leuschen, C. J., McAdoo, D., Manizade, S. S. and Raney, R. K.: Combined airborne laser and radar altimeter measurements over the Fram Strait in May 2002, Remote Sensing of Environment, 111, 182–194, doi:10.1016/j.rse.2007.02.037, 2007.

Girard-Ardhuin, F. and Ezraty, R.: Enhanced Arctic Sea Ice Drift Estimation Merging Radiometer and Scatterometer Data, IEEE Transactions on Geoscience and Remote Sensing, 50(7), 2639–2648, doi:10.1109/TGRS.2012.2184124, 2012.

Granskog, M. A., Rösel, A., Dodd, P. A., Divine, D., Gerland, S., Martma, T. and Leng, M. J.: Snow contribution to first-year and second-year Arctic sea ice mass balance north of Svalbard, Journal of Geophysical Research: Oceans, 122(3), 2539–2549, doi:10.1002/2016JC012398, 2017.

Grenfell, T. C. and Perovich, D. K.: Spectral albedos of sea ice and incident solar irradiance in the southern Beaufort Sea, J. Geophys. Res., 89(C3), 3573–3580, doi:10.1029/JC089iC03p03573, 1984.

Guerreiro, K., Fleury, S., Zakharova, E., Rémy, F. and Kouraev, A.: Potential for estimation of snow depth on Arctic sea ice from CryoSat-2 and SARAL/AltiKa missions, Remote Sensing of Environment, 186, 339–349, doi:10.1016/j.rse.2016.07.013, 2016.

Hines, K. M. and Bromwich, D. H.: Development and Testing of Polar Weather Research and Forecasting (WRF) Model. Part I: Greenland Ice Sheet Meteorology, Mon. Wea. Rev., 136(6), 1971–1989, doi:10.1175/2007MWR2112.1, 2008.

Holland, M. M., Bailey, D. A., Briegleb, B. P., Light, B. and Hunke, E.: Improved Sea Ice Shortwave Radiation Physics in CCSM4: The Impact of Melt Ponds and Aerosols on Arctic Sea Ice, J. Climate, 25(5), 1413–1430, doi:10.1175/JCLI-D-11-00078.1, 2011.

Holland, P. R. and Kimura, N.: Observed Concentration Budgets of Arctic and Antarctic Sea Ice, J. Climate, 29(14), 5241–5249, doi:10.1175/JCLI-D-16-0121.1, 2016.

Hunke, E. C., & Lipscomb, W. H.: CICE: the Los Alamos Sea Ice Model Documentation and Software User's Manual Version 4.1. T-3 Fluid Dynamics Group, Los Alamos National Laboratory, Los Alamos, 2010.

Ivanova, N., Pedersen, L. T., Tonboe, R. T., Kern, S., Heygster, G., Lavergne, T., Sørensen, A., Saldo, R., Dybkjær, G., Brucker, L. and Shokr, M.: Inter-comparison and evaluation of sea ice algorithms: towards further identification of challenges and optimal approach using passive microwave observations, The Cryosphere, 9(5), 1797–1817, doi:10.5194/tc-9-1797-2015, 2015.

Jakobson, E., Vihma, T., Palo, T., Jakobson, L., Keernik, H. and Jaagus, J.: Validation of atmospheric reanalyses over the central Arctic Ocean, Geophys. Res. Lett., 39(10), L10802, doi:10.1029/2012GL051591, 2012.

Kern, S., Khvorostovsky, K., Skourup, H., Rinne, E., Parsakhoo, Z. S., Djepa, V., Wadhams, P. and Sandven, S.: The impact of snow depth, snow density and ice density on sea ice thickness retrieval from satellite radar altimetry: results from the ESA-CCI Sea Ice ECV Project Round Robin Exercise, The Cryosphere, 9(1), 37–52, doi:10.5194/tc-9-37-2015, 2015.

Kimura, N., Nishimura, A., Tanaka, Y. and Yamaguchi, H.: Influence of winter sea-ice motion on summer ice cover in the Arctic, Polar Research, 32(1), 20193, doi:10.3402/polar.v32i0.20193, 2013.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K. and Takahashi, K.: The JRA-55 Reanalysis: General Specifications and Basic Characteristics, Journal of the Meteorological Society of Japan. Ser. II, 93(1), 5–48, doi:10.2151/jmsj.2015-001, 2015.

Koenig, L. S., Ivanoff, A., Alexander, P. M., MacGregor, J. A., Fettweis, X., Panzer, B., Paden, J. D., Forster, R. R., Das, I., McConnell, J. R., Tedesco, M., Leuschen, C. and Gogineni, P.: Annual Greenland accumulation rates (2009–2012) from airborne snow radar, The Cryosphere, 10(4), 1739–1752, doi:10.5194/tc-10-1739-2016, 2016.

Kurtz, N. T., Farrell, S. L., Studinger, M., Galin, N., Harbeck, J. P., Lindsay, R., Onana, V. D., Panzer, B. and Sonntag, J. G.: Sea ice thickness, freeboard, and snow depth products from Operation IceBridge airborne data, The Cryosphere, 7(4), 1035–1056, doi:10.5194/tc-7-1035-2013, 2013.

Kwok, R. and Cunningham, G. F.: ICESat over Arctic sea ice: Estimation of snow depth and ice thickness, J. Geophys. Res., 113(C8), C08010, doi:10.1029/2008JC004753, 2008.

Kwok, R. and Cunningham, G. F.: Variability of Arctic sea ice thickness and volume from CryoSat-2, Phil. Trans. Roy. Soc., Lond A, 373(2045), doi:10.1098/rsta.2014.0157, 2015.

Kwok, R. and Maksym, T.: Snow depth of the Weddell and Bellingshausen sea ice covers from IceBridge surveys in 2010 and 2011: An examination, J. Geophys. Res. Oceans, 119(7), 4141–4167, doi:10.1002/2014JC009943, 2014.

Kwok, R. and Markus, T.: Potential basin-scale estimates of Arctic snow depth with sea ice freeboards from CryoSat-2 and ICESat-2: An exploratory analysis, Advances in Space Research, doi:10.1016/j.asr.2017.09.007, 2017.

Kwok, R., Kurtz, N. T., Brucker, L., Ivanoff, A., Newman, T., Farrell, S. L., King, J., Howell, S., Webster, M. A., Paden, J., Leuschen, C., Macgregor, J. A., Richter-Menge, J., Harbeck, J. and Tschudi, M.: Inter-comparison of snow depth retrievals over Arctic sea ice from radar data acquired by Operation IceBridge, The Cryosphere Discuss., 2017, 1–37, doi:10.5194/tc-2017-103, 2017.

Lavergne, T., Eastwood, S., Teffah, Z., Schyberg, H. and Breivik, L.-A.: Sea ice motion from low-resolution satellite sensors: An alternative method and its validation in the Arctic, J. Geophys. Res., 115(C10), C10032, doi:10.1029/2009JC005958, 2010.

Laxon, S. W., Giles, K. A., Ridout, A. L., Wingham, D. J., Willatt, R., Cullen, R., Kwok, R., Schweiger, A., Zhang, J., Haas, C., Hendricks, S., Krishfield, R., Kurtz, N., Farrell, S. and Davidson, M.: CryoSat-2 estimates of Arctic sea ice thickness and volume, Geophys. Res. Lett., 40(4), 732–737, doi:10.1002/grl.50193, 2013.

Lecomte, O., Fichefet, T., Flocco, D., Schroeder, D. and Vancoppenolle, M.: Interactions between wind-blown snow redistribution and melt ponds in a coupled ocean–sea ice model, Ocean Modelling, 87, 67–80, doi:10.1016/j.ocemod.2014.12.003, 2015.

Leonard, K. C. and Maksym, T.: The importance of wind-blown snow redistribution to snow accumulation on Bellingshausen Sea ice, Annals of Glaciology, 52(57), 271–278, doi:10.3189/172756411795931651, 2011.

Lindsay, R., Wensnahan, M., Schweiger, A. and Zhang, J.: Evaluation of Seven Different Atmospheric Reanalysis Products in the Arctic*, J. Climate, 27(7), 2588–2606, doi:10.1175/JCLI-D-13-00014.1, 2014.

Liston, G. E., Polashenski, C., Rösel, A., Itkin, P., King, J., Merkouriadi, I. and Haapala, J.: A Distributed Snow-Evolution Model for Sea-Ice Applications (SnowModel), Journal of Geophysical Research: Oceans, 123(5), 3786–3810, doi:10.1002/2017JC013706, 2018.

Maass, N., Kaleschke, L., Tian-Kunze, X. and Tonboe, R. T.: Snow thickness retrieval from L-band brightness temperatures: a model comparison, Annals of Glaciology, 56(69), 9–17, doi:10.3189/2015AoG69A886, 2015.

Maksym, T. and Markus, T.: Antarctic sea ice thickness and snow-to-ice conversion from atmospheric reanalysis and passive microwave snow depth, J. Geophys. Res., 113(C2), C02S12, doi:10.1029/2006JC004085, 2008.

Markus, T. and Cavalieri, D. J.: Snow Depth Distribution Over Sea Ice in the Southern Ocean from Satellite Passive Microwave Data, in Antarctic Sea Ice: Physical Processes, Interactions and Variability, edited by rtin O. Jeffries, pp. 19–39, American Geophysical Union., 1998.

Markus, T., Neumann, T., Martino, A., Abdalati, W., Brunt, K., Csatho, B., Farrell, S., Fricker, H., Gardner, A., Harding, D., Jasinski, M., Kwok, R., Magruder, L., Lubin, D., Luthcke, S., Morison, J., Nelson, R., Neuenschwander, A., Palm, S., Popescu, S., Shum, C., Schutz, B. E., Smith, B., Yang, Y. and Zwally, J.: The Ice, Cloud, and land Elevation Satellite-2 (ICESat-2): Science requirements, concept, and implementation, Remote Sensing of Environment, 190, 260–273, doi:10.1016/j.rse.2016.12.029, 2017.

Markus, T., Stroeve, J. C. and Miller, J.: Recent changes in Arctic sea ice melt onset, freezeup, and melt season length, J. Geophys. Res., 114, doi:10.1029/2009JC005436, 2009.

Massom, R. A., Eicken, H., Hass, C., Jeffries, M. O., Drinkwater, M. R., Sturm, M., Worby, A. P., Wu, X., Lytle, V. I., Ushio, S., Morris, K., Reid, P. A., Warren, S. G. and Allison, I.: Snow on Antarctic sea ice, Rev. Geophys., 39(3), 413–445, doi:10.1029/2000RG000085, 2001.

Maykut, G. A. and Untersteiner, N.: Some results from a time-dependent thermodynamic model of sea ice, J. Geophys. Res., 76(6), 1550–1575, doi:10.1029/JC076i006p01550, 1971.

Meier, W. N.: Comparison of passive microwave ice concentration algorithm retrievals with AVHRR imagery in arctic peripheral seas, IEEE Transactions on Geoscience and Remote Sensing, 43(6), 1324–1337, doi:10.1109/TGRS.2005.846151, 2005.

Merkouriadi, I., Gallet, J.-C., Graham, R. M., Liston, G. E., Polashenski, C., Rösel, A. and Gerland, S.: Winter snow conditions on Arctic sea ice north of Svalbard during the Norwegian young sea ICE (N-ICE2015) expedition, J. Geophys. Res. Atmos., 122(20), 2017JD026753, doi:10.1002/2017JD026753, 2017.

Panzer, B., Gomez-Garcia, D., Leuschen, C., Paden, J., Rodriguez-Morales, F., Patel, A., Markus, T., Holt, B. and Gogineni, P.: An ultra-wideband, microwave radar for measuring snow thickness on sea ice and mapping near-surface internal layers in polar firn, Journal of Glaciology, 59(214), 244–254, doi:10.3189/2013JoG12J128, 2013.

Perovich, D., Polashenski, C., Arntsen, A. and Stwertka, C.: Anatomy of a late spring snowfall on sea ice, Geophys. Res. Lett., 44(6), 2016GL071470, doi:10.1002/2016GL071470, 2017.

Perovich, D. K., Grenfell, T. C., Light, B. and Hobbs, P. V.: Seasonal evolution of the albedo of multiyear Arctic sea ice, J. Geophys. Res., 107(C10), 8044, doi:10.1029/2000JC000438, 2002.

Petrich C., Eicken H., Polashenski C. M., Sturm M., Harbeck J. P., Perovich D. K. and Finnegan D. C.: Snow dunes: A controlling factor of melt pond distribution on Arctic sea ice, J. Geophys. Res, 117(C9), doi:10.1029/2012JC008192, 2012.

Petty, A. A., Stroeve, J. C., Holland, P. R., Boisvert, L. N., Bliss, A. C., Kimura, N. and Meier, W. N.: The Arctic sea ice cover of 2016: a year of record-low highs and higher-than-expected lows, The Cryosphere, 12(2), 433–452, doi:10.5194/tc-12-433-2018, 2018.

Pomeroy, J. W., Marsh, P. and Gray, D. M.: Application of a distributed blowing snow model to the Arctic, Hydrol. Process., 11(11), 1451–1464, doi:10.1002/(SICI)1099-1085(199709)11:11<1451::AID-HYP449>3.0.CO;2-Q, 1997.

Radionov, V. F., Bryazgin, N. N. and Alexandrov, E. I.: The Snow Cover of the Arctic Basin, Technical Report, APL-UW TR 9701, 1997.

Rienecker, M. M., Suarez, M. J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M. G., Schubert, S. D., Takacs, L., Kim, G.-K., Bloom, S., Chen, J., Collins, D., Conaty, A., da Silva, A., Gu, W., Joiner, J., Koster, R. D., Lucchesi, R., Molod, A., Owens, T., Pawson, S., Pegion, P., Redder, C. R., Reichle, R., Robertson, F. R., Ruddick, A. G., Sienkiewicz, M. and Woollen, J.: MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications, J. Climate, 24(14), 3624–3648, doi:10.1175/JCLI-D-11-00015.1, 2011.

Serreze, M. C. and Stroeve, J.: Arctic sea ice trends, variability and implications for seasonal ice forecasting, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 373(2045), 2015.

Serreze, M. C., Barrett, A. P., Slater, A. G., Woodgate, R. A., Aagaard, K., Lammers, R. B., Steele, M., Moritz, R., Meredith, M. and Lee, C. M.: The large-scale freshwater cycle of the Arctic, J. Geophys. Res., 111(C11), C11010, doi:10.1029/2005JC003424, 2006.

Sturm, M.: Field Techniques for snow observations on sea ice, in Field techniques for sea ice research, H. Eicken, R. R. Gradinger, M. Salganek, K. Shirasawa, D. K. Perovich and M. Leppäranta (Eds.), University of Alaska Press, Fairbanks., 2009.

Sturm, M. and Massom, R. A.: Snow in the sea ice system: friend or foe?, in Sea Ice, edited by D. N. Thomas, pp. 65–109, John Wiley & Sons, Ltd., 2017.

Sturm, M. and Stuefer, S.: Wind-blown flux rates derived from drifts at arctic snow fences, Journal of Glaciology, 59(213), 21–34, doi:10.3189/2013JoG12J110, 2013.

Sturm, M., Perovich, D. K. and Holmgren, J.: Thermal conductivity and heat transfer through the snow on the ice of the Beaufort Sea, J. Geophys. Res., 107(C10), 8043, doi:10.1029/2000JC000409, 2002.

Sumata, H., Lavergne, T., Girard-Ardhuin, F., Kimura, N., Tschudi, M. A., Kauker, F., Karcher, M. and Rudiger, G.: An intercomparison of Arctic ice drift products to deduce uncertainty estimates, J. Geophys. Res. Oceans, 119, doi:10.1002/2013JC009724, 2014.

Szanyi, S., Lukovich, J. V., Barber, D. G. and Haller, G.: Persistent artifacts in the NSIDC ice motion dataset and their implications for analysis, Geophys. Res. Lett., 2016GL069799, doi:10.1002/2016GL069799, 2016.

Trujillo, E., Leonard, K., Maksym, T. and Lehning, M.: Changes in snow distribution and surface topography following a snowstorm on Antarctic sea ice, J. Geophys. Res. Earth Surf., 2016JF003893, doi:10.1002/2016JF003893, 2016.

Tschudi, M., Fowler, C., Maslanik, J. and Stewart, J. S.: Polar Pathfinder Daily 25 km EASE-Grid Sea Ice Motion Vectors, Version 3. Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center., , doi:0.5067/O57VAIT2AYYY, 2016.

Warren, S. G.: Optical properties of snow, Rev. Geophys., 20(1), 67–89, doi:10.1029/RG020i001p00067, 1982.

Warren, S. G., Rigor, I. G., Untersteiner, N., Radionov, V. F., Bryazgin, N. N., Aleksandrov, Y. I. and Colony, R.: Snow Depth on Arctic Sea Ice, J. Climate, 12(6), 1814–1829, doi:10.1175/1520-0442(1999)012<1814:SDOASI>2.0.CO;2, 1999.

Webster, M. A., Rigor, I. G., Nghiem, S. V., Kurtz, N. T., Farrell, S. L., Perovich, D. K. and Sturm, M.: Interdecadal changes in snow depth on Arctic sea ice, J. Geophys. Res. Oceans, 119(8), 5395–5406, doi:10.1002/2014JC009985, 2014.

Worby, A. P., Markus, T., Steer, A. D., Lytle, V. I. and Massom, R. A.: Evaluation of AMSR-E snow depth product over East Antarctic sea ice using in situ measurements and aerial photography, J. Geophys. Res., 113(C5), C05S94, doi:10.1029/2007JC004181, 2008a.